

20 May 1982

Good start on strategic arms

President Ronald Reagan's proposals for a new bilateral negotiation on strategic arms control are welcome. Will the Soviet Union listen? And will the Administration follow through?

President Reagan appeared to take a stride towards serious strategic arms control on 9 May when he made his first strategic arms control proposal to the Soviet Union and called for deep cuts in each side's arsenals. While the proposal deserves applause as a first step, it is like the proverbial horse that has been led to water but has not yet drunk. The Soviet Union has yet to respond. And unless it is followed by clear, realistic and consistent effort by the Administration in the coming months, even years, the progress Reagan has made will be undone.

The proposal, transmitted to the Soviet Union in detailed form at the same time as the President outlined it in a speech at his *alma mater*, Eureka College in Illinois, focuses on reducing the Soviet advantage in long-range intercontinental ballistic missiles (ICBMs) by asking that each side should reduce to 5,000 the number of its ballistic missile warheads, no more than half of which could be on land-based ICBMs. At the latest count, the Soviet Union may have 6,300 ballistic missile warheads, of which 5,100 are on land-based ICBMs. The United States has 7,500 ballistic missile warheads, of which only 2,154 are on land-based ICBMs. One effect of the proposal is that the Soviet Union would have to reduce its land-based ICBM warheads from 5,100 to 2,500, and so forgo all or most of its force of SS-18 giant "heavy" missiles, which could have up to 3,000 high-precision warheads and which could now, in theory, knock out most of the US force of land-based ICBMs in a single blow. The United States, which invests a smaller fraction of its strategic forces in land-based ICBMs, is at present below the 2,500 warhead ceiling in the President's proposal and could comply merely by retiring some of its older ICBM-carrying submarines without sacrificing much in force capability. So much for phase one.

The President also proposed a second phase of negotiations in which both sides would strive for equality by another measure, ballistic missile throw-weight, "at less than current American levels". Throw-weight is the weight of a missile's payload including warhead. This notion has grown out of an argument (backed by Eugene V. Rostow, director of the Arms Control and Disarmament Agency, and by Defense Secretary Caspar Weinberger) that an aggregate index of throw-weight would be a better basis for measuring current and future force levels than merely counting holes in the ground or launch tubes on submarines, as in Salt II. The President apparently resisted using this formula for the first phase of negotiations because the joint chiefs of staff felt that weight-counts are less verifiable than actual warheads. Rostow now says that the connection between phase one and phase two will depend on the Soviet Union's reaction and counter-suggestions, but that phase one alone would achieve substantial reductions in throw-weight on the Soviet side, the area where its forces have an advantage.

So far, the plan has only been sketched in public, and the Soviet side has just received it. More details may not be forthcoming for some time, but the plan has some obvious weaknesses and some curious strengths. One weakness is that the proposal does not include nuclear warheads carried by aircraft in its warhead ceiling of 5,000. Instead, the proposal is said to put a separate ceiling on bombers, limiting each side to between 350 and 400. On the US side, this would include the long-range older B-52 bombers and the new B-1 but also the intermediate range FB-111s based in Europe, armed with nuclear weapons and capable of striking at

targets in the Soviet Union. The Soviet Union has long insisted that the FB-111s should be counted as strategic weapons, a view which the United States has rejected in the past. In return for including the FB-111s under the bomber ceiling, the United States would now insist that the Soviet Union includes the "Backfire" bomber under its bomber ceiling. The range and deployment of the Backfire bomber was a contentious issue in Salt II — and its exclusion from the Salt II ceilings was one reason why that agreement was never ratified. In this connection, domestically, the Reagan plan could give the Administration a respectable way to back out of building large numbers of MX missiles, which have been touted as — but not proved to be — a way to reduce the vulnerability of land-based ICBMs, although the obvious way for the United States to reduce its total number of ballistic missile warheads from 7,200 to 5,000 would be to scrap some older submarines, another could be to cancel part of the MX programme. The Administration is already embarrassed that its promise a year and a half ago to set the MX issue straight forthwith has not yet borne fruit.

The Reagan plan therefore deserves a fair wind. It is to be hoped that the Soviet Union will not reject it outright, as it rejected President Carter's proposal for deep cuts in missile forces in 1977. But advocates of a nuclear freeze in the United States (see *Nature* 29 April, p.790) and Europe should not delude themselves that the President has taken a step towards their version of arms control. A curious fact about the politics of arms control in the United States is that both the left and the right favour deep cuts in nuclear forces, but for different reasons. The left favours cuts (and the extreme left favours unilateral cuts) as part of the continuing process of negotiation. The right, in the United States, favours proposals for deep cuts on the grounds that they are a way of limiting Soviet striking power substantially — and that they also challenge Soviet protestations of good intentions on arms control. At the worst, it is always possible to say "See, we told you so. They don't mean what they say".

As it happens, it is not yet clear how sincere even the Reagan Administration is about arms control, no matter what the President said in Eureka. In their 18-month partnership, Secretary of State Alexander Haig and Secretary of Defense Caspar Weinberger have shown that they have radically different views of the US-Soviet military balance. As a result of these and other differences, often aired publicly, the Administration has yet to convince countries abroad (and even itself) where it stands on some key issues. Yet what is critical to the success of the strategic arms negotiations is not this opening flourish but whether, in following up, the Administration can speak with one voice and settle down to serious talks, taking Soviet rhetoric in its stride and keeping policy on a single course, rather than zigzagging and confusing the Soviet leadership. In short, it remains to be seen whether the horse will take a drink.

Unfortunately it will be several months before these issues can be clarified. Mr Brezhnev, having had the Soviet Union complain that President Reagan was "warmongering" by choosing to attend the NATO meeting planned for June and not the opening session of the Special Session on disarmament of the United Nations Assembly, has now ducked the United States offer (or challenge) of a meeting on that occasion. Instead, the two men may now meet on neutral ground sometime in October. The delay



is probably to be welcomed. It will give the US Administration a chance to put its act together, and the Soviet Union a chance to resolve the delicate problem of continuity in strategic arms negotiations during the period when Mr Brezhnev will be succeeded by somebody else. And the delay will also be a chance to decide how the negotiations now proposed should be linked with those formally under way, but almost stalled, on the limitation of nuclear weapons in Europe. The United States has always intended that the two processes should be merged, as it is sensible that they should be. For only then will President Reagan's "zero option" seem secure. By October, the Soviet Union should be ready to soften the negotiating position it has repeated since last November — that it will withdraw some of its mobile missiles if the United States will agree not to introduce the Pershing II and cruise missiles.

French new world?

France wants next month's economic summit to talk technology. It should not expect too much.

The summer now begun in the Northern Hemisphere will be metaphorically long and hot. Several important issues fall to be decided, or at least to be given an airing. Strategic arms control (see above) is one. The South Atlantic (where it will be winter) is another. Next month, there will also be the Special Session of the United Nations on Disarmament in New York. Before then, President Reagan will have made the rounds of several West European capitals and will have sat in on two crucial if lesser summit-meetings — one (in Bonn) of the North Atlantic Treaty Organization (NATO) and another (at Versailles) of the members of the sub-set of the Organization for Economic Cooperation and Development (OECD) considered to deserve a seat at gatherings where senior politicians contemplate the future of the world. President François Mitterrand of France, who will not be at Bonn because France does not belong to NATO, has characteristically done his best to ensure that the Versailles summit, under the wing of OECD, will be enlivening. He has hatched a plan that the economic summit should begin with a discussion of governments' responsibilities for the development of technology. If he were not the host, the *rentier* in the public cause of the splendid palace at which the meeting will take place, it is improbable that his subject would have reached the head of the agenda. But the subject cannot be the emollient of fellow feeling that he calculated it to be.

President Mitterrand may in retrospect be thought to have captured the allegiance of those who voted for him a year ago by his simplicity; he seems to believe in what he says. This phenomenal habit appears also to be the reason why he has also won the affection, even respect, of European politicians ideologically at odds with him. But President Mitterrand also believes in the future — a future fashioned by technology. Why else would he have given such encouragement to his buoyant colleague, M. Jean-Pierre Chevènement, minister for research and technology? Together, they appear to be convinced that the future of France can be assured by the intelligent direction and the generous support of research and development. They may well be right. But President Mitterrand is also now bent on propagating this hopeful message in a wider circle, among France's industrial partners and competitors. The objectives may again be admirable, but it is important that President Mitterrand should understand why his advocacy next month at Versailles of the case that science and technology are a means of making up for economic recession will cause many of those present to shuffle uneasily in their seats. For the French argument, while half right, is three-quarters wrong.

The case that science and technology are too much neglected by those who look for a way out of present economic troubles is, of course, beyond dispute. If the present recession is caused by the fivefold increase of the world price of petroleum in the past decade, then something like a ten per cent increase of productivity in the oil-consuming states would suffice to ensure that nobody

would notice. That, of course, is well within the immediate grasp of people in research and development. And a fivefold increase of productivity, by no means out of court, would ensure that the cost of buying oil would at some point in the future seem as unimportant as in the early 1960s, but with the difference that average prosperity would be five times greater. So, by putting money into research and development, the governments of industrialized states can realistically hope to work their way towards a more stable world, one in which not merely their electors but the people who live elsewhere will have a sense of being better off. Would that President Mitterrand's colleagues at next month's meeting at Versailles more openly shared his view.

The weakness of the French case is that it is almost certain to be an exaggeration. For at least the past two decades, successive French governments have worked deliberately at the improvement of science and technology. And not before time. The old republic was uncongenial for research. President de Gaulle, concerned though he may have been with the paramountcy of the identity of France, and of its place in the world, nevertheless provided an opportunity for those who sought to remedy the earlier neglect of the universities and of research. It tends, however, to be forgotten that even the young turks who took up that challenge made several disastrous mistakes. While the development of the petrochemicals industry in the early 1960s was an immense success, a proof that technology can spell prosperity, the nuclear power industry of that time was swiftly overtaken by events (and by the superior designs put on offer by United States manufacturers) while the efforts of those early governments to build a domestic computer industry were no more successful than later attempts have been. The still new government of France has inherited a much stronger base for the conduct of research and development than any of its predecessors; the universities are mostly in good shape, as are the grant-making agencies. But the government seems not to have learned from the hard experience of earlier decades. Its political persuasions, indeed, tell it that it need not learn. Yet mistakes will happen, and governments are especially prone to them.

This is why the unexpected French agenda item for the Versailles summit meeting will not be the uncontroversial matter that President Mitterrand seems to hope. While some among the governments of the United Kingdom, the United States and West Germany may have ideological reasons for believing that too much government intervention in the shaping of the pattern of technology is mistaken, each of them also has vivid illustrations of how such efforts can go wrong. Sir Harold Wilson's technological revolution (1974) is only the most memorable error. Even the Japanese experience is not the proof that sufficiently clever governments can act wisely that the French will be hoping to claim at Versailles. For while the Tokyo government's support of successful manufacturers is the envy of manufacturers elsewhere (and properly a cause for complaint by luckless importers in other countries), nobody suggests that the government of Japan plans in advance whether next year's wonder product will be a new hi-fi set or a new space invaders game. The principle seems to be that the market will define what technology should do — if the market is given half a chance.

The French, in their present mood, are unconvinced. And it is certainly the case that if President Mitterrand could win the agreement of other industrialized nations to some coordinated plan for the development of technology, it would be possible to create both rising prosperity and a proof that the technological prophecy had come true. The agreement will not be forthcoming, but the proof would in any case be false. For to ordinary people, governments' electors, what matters about technology is not that it should be successful but that it should be wanted. Even the belief (to which the French government holds) that science and technology should be harnessed to the needs of development overseas is likely to run foul of the now well-established truth that developing countries do not like being told what is good for them. Sooner or later, Frenchmen will come to hold the same opinion if President Mitterrand pushes his view of the management of technology down their throats.

US sours on foreign graduates

Job prospects clouded by immigration bill

Washington

US high technology industries and scientific organizations are just waking up to a little-noted, devastating change in US immigration laws which, if enacted, could change the face of Silicon Valley and university engineering faculties. The proposed change would prohibit foreign students from taking jobs in the United States on graduation. Instead, they would have to go home for two years before they could apply to work in the United States.

Under present law, they may apply for a change in visa status without leaving the country, and often do so.

The new provision may well become law before Congress adjourns. Identically worded provisions are in both the House and Senate versions of the bill. Neither has yet emerged from committee, but the provisions are unlikely to change before Congress acts.

Nobody knows how many students would be affected partly because many are either in the country illegally or do not report their work status.

At Stanford, for example, 50 per cent of all students studying for a PhD degree are foreign. Nationally, in 1980, 20.4 per cent of all science and engineering graduate students were foreign. In engineering, one-third of all PhD recipients are not US citizens. Engineering is the most popular field for the 311,000 foreign students, graduate and undergraduate, estimated to be in the United States (see Table).

Civilian high technology companies have come to rely increasingly on newly qualified engineers and scientists who are not US citizens, says Pat Hubbard of the American Electronics Association, because the US defence build-up has caused defence industries to hive off many of the American graduates. William Cagney of the National Foreign Trade Council which represents 650 major companies adds: "When DuPont goes looking for a PhD in chemical engineering, it more often than not hires a foreigner; there simply aren't that many Americans to go around. These people are vital to US industry." Both trade groups are opposing the proposed change in the law, as is the Semiconductor Industry Association, representing many Silicon Valley firms.

The provision in question is Section 212 of a bill that changes the Immigration and Nationality Act, the umbrella law that governs US immigration policies. Reform of the act has long been a public issue in the

United States, where many thousands of Mexicans cross the southern border to work on ranches, illegally, and where the plight of the Cuban and Haitian refugees drew world attention to the problems of assimilating still more people from abroad. A blue-ribbon study was made by the Select Commission on Immigration, headed by Father Theodore Hesbergh of Notre Dame. The Reagan Administration proposed its own views of how to tighten US immigration laws, but the bill that has received the most support is that proposed by Romano L. Mazzoli (Democrat, Kentucky) in the House and Alan K. Simpson (Republican, Wyoming) in Senate. That part of the bill involving foreign students did not appear in Father Hesbergh's recommendations nor in the

Administration bill, and so went unnoticed until a few weeks ago.

At present, most foreign students in the United States have visas allowing them to stay as long as they are studying, but they are not allowed to work. On graduation, they may apply for permanent resident status — a step towards citizenship — if they have a job offer. They usually remain in the United States (working) until this change occurs. A minority of foreign students are in the United States on so-called J visas, having been sent by their government or on exchange programmes, and they must return home on completion of their studies. Under Section 212, all students would have the equivalent of J status.

A House staff member explained that far

Setback for Chevènement

The second constitutional airing of the prospective science and technology bill (*loi*), on which the French science boom depends, has led to a bloodbath. For the Senate, the conservative French upper house, last week voted amendments which amputate most of its effective clauses.

The law, in the tattered form approved by the Senate, will now go to the National Assembly (the lower house), probably in June. Although the government has the majority there, the law will have to go through a lengthy process in which clauses deleted by Senate are reintroduced as amendments, adding to an already packed parliamentary session.

The Senate is an indirectly elected body, a third of its members being appointed each year on the basis of regional (departmental) elections. Electors consist of mayors and other such dignitaries, and amount to some 500 from each department. The Senate remains firmly conservative despite last year's election of the socialist Mitterrand government, and so is a source of constant irritation. But this is the first time it has behaved quite as badly towards government legislation.

The government explanation is that the science and technology bill gave the opposition an excellent opportunity to put a major dent in the plans of a senior minister, Jean-Pierre Chevènement, who is also the leader of the most left-wing element of the socialist party, the Ceres group. Ceres is of an intellectual and loosely Marxist tendency, and a favourite target of the conservatives. The fact that science and technology were involved in the Senate debacle is incidental.

Chevènement wished to devolve power to the regions, setting up regional councils for science and technology with their own budgets. The senators wished to "put off" that reform.

Chevènement wanted to open up

research institutions such as the Centre National de la Recherche Scientifique (CNRS) to industry, by creating a new type of legal entity through which for example, CNRS could make profits in joint ventures. The senators would have none of it.

And, most of all, Chevènement wanted a law that would define future budgets: 17.8 per cent extra each year in real terms over the next three years. "Too imprecise" said the senators, who feel that such a broad commitment would give Chevènement too much power. In fact, the minister is now discovering that a commitment to a figure on paper — even in a parliamentary bill —

"L'ÉTAT, C'EST MA LOI"



is not a firm promise. Discussions are now going on at cabinet level for the 1983 budget, and Chevènement is having to fight as hard this year as he did last.

Chevènement was impatient with the Senate's reaction to his well-laid plans. He described the Senate majority as "reactionary" and "cut off from the people". But he now faces a long struggle against the conservatives and against those of his cabinet colleagues who resent his scientific imperialism. **Robert Walgate**

too many foreigners arrive in the United States saying they are going to study, but really plan to stay afterwards and work. This is a "backdoor immigration policy", he said, which must be reformed as part of a general tightening of US immigration.

The provision may have originated with an engineering activist, Irwin Feerst. Feerst says there was no consideration of a change in the old law regarding students, until he sent out a special issue of an independent newsletter he publishes to engineers around the country. Three hundred responses he received, he says, indicated that professional US engineers want to "throw the foreign students out". Feerst spoke to Senator Simpson and testified to this effect in December. According to Feerst, there is

Nationality of foreign students in the United States 1980-81

Regions	Selected sub-totals	Totals
Africa		38,180
Nigeria	17,350	
Europe		25,330
UK	4,440	
Greece	3,750	
FRG	3,310	
France	2,570	
Eastern Europe + USSR	1,670	
Latin America		49,810
Middle East		84,710
Iran	47,550	
Saudi Arabia	10,440	
North America		14,790
Oceania		4,180
South and East Asia		94,640
Taiwan	19,460	
Japan	13,500	
India	9,250	
		311,640

Source: Institute of International Education

no shortage of engineers in the United States, only a "cabal" of academic engineers and corporate executives who are publicizing the alleged shortage so they can hire foreign graduates at lower pay.

However, the Institute of Electrical and Electronic Engineers (IEEE), the engineers' umbrella society in the United States (where Feerst has often run as an "alternative" candidate for president) does not favour the change in Section 212 as it is now written. Richard J. Gowen, chairman of IEEE's manpower task force and a candidate for president of IEEE, says that foreign engineers are being offered jobs at salaries that may be as little as 25 per cent of the salaries paid to US-born engineers. Gowen wants the proposed Section 212 to be changed to allow only "professional" hiring of foreign graduates. That is, a foreign graduate would have to go home unless his prospective employer certifies that he will be paid at least 75 per cent of what a US citizen would be paid in the job.

So far, Congress has been mainly concerned with getting the immigration reform bill through, and seems little disposed to tinker with the minor provisions. On the other hand, organizations like IEEE are becoming very active on the issue, as they begin to realize its implications.

Deborah Shapley

EEC budget

Windfall ahead

Brussels

An unexpected shortfall of £280 million (500 million European Currency Units) in the European Economic Community's agricultural expenditure has resulted in an extra £19.6 million (35 million ECUs) becoming available for the European Community's research programmes in 1982. The strength of the dollar, favourable prices on the world market for agricultural goods and good weather have all helped to reduce the EEC's agricultural subsidies. Most of the extra money will be allocated to the joint research centres and in particular to nuclear safety research, but around £3.7 million will go to the indirect action programmes.

The budgetary revision highlights the degree to which the European Commission's ability to meet its goal of revitalizing scientific research and development in Europe, and thus create new jobs for scientists, is linked with the long-standing quarrel over agricultural prices, the British budgetary contribution, the reform of the Community's budget and the Common Agricultural Policy and even the Falklands crisis. If an agreement could be reached on lower agricultural prices for this year, as the British are demanding, an even greater sum could be set free for other areas of expenditure. The cooperation shown by the other member states over the Falklands crisis has now made it more difficult for Britain to push for budgetary reform. But if Britain achieves its objectives in the Council meetings this week, not only will its contributions be reduced, but it should lead to greater Community outlays in other areas including research.

The Commission's preliminary draft budget for 1983 reflects the hope that the other nine member states will agree to shift the emphasis of Community spending. Thus, the text emphasizes a "significant reinforcement of financial resources for energy policy, innovation and research and development". This includes increasing expenditure on energy research by 120 per cent although the total sum for payment appropriations will still be small, at £58.4 million.

Most of the extra money for this year will go towards the Supersara project on reactor safety at the joint research centre at Ispra in Italy. An extra £4.2 million will be needed this year as the project is overshooting its budget and will probably continue doing so until 1990 to the tune of £168 million.

The entire programme for the joint research centres is, in fact, now being reviewed and some changes are certain to be reflected in the 1984-87 research programme. Under consideration is a temporary increase in staff by 161, to provide replacements for scientists expected to retire in the next few years.

Apart from streamlining research on nuclear safety, the handling of radioactive wastes and the control of fissile materials, a new institute for developing countries is planned at Ispra for training in energy planning, new energies, remote sensing techniques and the inventoring of resources. The Commission is also hoping to increase staff for research into solar energy, fusion, the rational use of energy and the study of high temperature materials.

Again, the success of these proposals will depend on the attitude of the member states, who will be influenced by the amount of money in the budget left over after provisions have been made for agricultural subsidies.

Jasper Becker

Polish arrest

Expel and detain

The expulsion from Poland, last week, of two US diplomats and the arrest of Dr Ryszard Herczynski bodes ill for the resumption of normal academic exchanges between Poland and the West. The two diplomats, Scientific Attaché John William Zerolis and First Secretary for Cultural Affairs, James Daniel Howard, found in Dr Herczynski's flat, were accused by the Polish ministry of foreign affairs of "pursuing activity conflicting with their diplomatic status". Their presence in the flat appears, however, to have been entirely in the line of duty; they had gone there to confer with Dr Herczynski and with Professor Wladyslaw Fiszdon, a former pro-rector of Warsaw University, on the forthcoming joint US-Polish symposium on fluid mechanics.

Dr Herczynski, a mathematician specializing in fluid dynamics, is employed at the Polish Academy of Sciences' Institute for Fundamental Problems of Technology. Although now accused of having been "one of the inspirers of activity contrary to our *raison d'être* in the scientific milieu", he has never been associated with the dissident movement. In autumn 1980, however, he founded the "Society of the friends of science" — a semi-popular organization based on the then current principles of the liberalization of learning.

It was presumably for this reason that, during the night of 12-13 December 1981, he was taken into custody and interned for some two weeks. Although the authorities now claim that before being released, he signed an undertaking to cease such activities, Dr Herczynski's friends insist that not only did he never sign such an undertaking, but that at the time of his release there had been no mention of any such document.

The arrest of Dr Herczynski was, according to official sources, effected as he handed Mr Zerolis a packet of material "damaging to the interests of the Polish People's Republic", apparently com-

prising a personal letter to his son, at present studying in the United States, an unidentified leaflet and an anonymous "Code of conduct during this testing time" addressed to Polish academics. Dr Herczynski now faces trial before a summary court. **Vera Rich**

NATO civil research

More wanted

Applicants for NATO (North Atlantic Treaty Organization) fellowships will have less chance of success this year than ever before. Applications have risen by 30 per cent, while the number of fellowships (about 800) remains constant. The NATO science committee will have to apply "new criteria" to make selections, a spokesman said last week. One possibility is that group applicants will be favoured over individuals.

NATO staff pinpoint four reasons for the increase in applications — financial difficulties among the 15 member countries, an increase in the number of papers now citing NATO as a supporting agency, a deliberate "willingness" on the part of NATO to expand the programme and a policy of greater visibility, including advertising in *Nature*.

The willingness to expand, however, is restricted to the NATO Science Committee, headed by Frenchman Professor Robert Chabbal (at present NATO's Assistant Secretary-General for Scientific and Environmental Affairs). The Civil Budget Committee, from which Chabbal draws his funds, is not so willing. To cope with the increase in applications for fellowships and for summer-school sponsorship, and for increased travel costs, the committee would have to increase its budget next year by 25–30 per cent in real terms, to \$23–24 million. In fact, it may be lucky to get 15 per cent extra, just enough to cover the depreciation of the Belgian franc.

Pressing his case, Chabbal claimed last week that the NATQ civil science programme (which completely avoids military research) is substantial and important. It accounts for half of all summer-school and training fellowships. Schools such as the Ettore Majorana at Erice, Sicily, and Les Houches in the Alps, get 60 per cent of their money from NATO, said Chabbal.

Meanwhile the committee will press ahead with new plans. It runs advisory panels which help to provide seed money for communications in new fields, largely by establishing "advanced workshops", and this year it will create two new panels: one on global transport mechanisms (in the atmosphere, ocean and mantle) and one on the selective activation of molecules (for example by laser). These panels would be expected to launch six workshops a year for a maximum of five years.

The committee is also experimenting with links between an industry in one

country and a university in another, in a programme dubbed the "double jump". Finance will be *à la carte* — only interested countries need support it. So far only two such fellowships have been organized, but many more are planned — Dr Mario di Lullo, organizer of the "double jump" programme, believes that it will not run into the protectionist difficulties that sometimes face the European Commission — that one nation's industry does not reveal its secrets to nationals of another. **Robert Walgate**

British biotechnology

Out of the blue

In an unusual move, the British University Grants Committee is earmarking part of its annual budget to develop a specific topic — biotechnology. The committee plans that £800,000 will be spent in each of the next three academic years on fostering biotechnology in a handful of universities. The committee had previously been reluctant to earmark grants, preferring that universities should spend their income as they wished, relying on the research councils to encourage centres in particular topics by means of research grants.

The scale of the recent budget cuts seems to have prompted a change of heart. The committee, worried that universities may pare all their activities rather than cut them selectively, clearly hopes that earmarked grants will make the future pattern of university research more pointed. The £800,000 set aside for biotechnology will be taken from the money reserved for restructuring the reduced university system which in the next academic year (1982–83) will be £50 million. Most of that sum is expected to be spent on payments to redundant academics, leaving uncertain the amount available for fostering priorities.

So far, three centres — at University College London, the University of Birmingham and the University of Manchester Institute of Science and Technology — have been awarded annual grants of £100,000 each to develop biotechnology. Five other centres are expected to receive similar grants soon. The money will be paid as a separate item in each of the next three years, after which it will be incorporated in the recipients' recurrent grants.

The grants committee says that the recipients must decide for themselves how to spend their extra money. Nevertheless, it expects them to forge closer links with industry, chiefly by encouraging the process engineering side of biotechnology, to develop postgraduate rather than undergraduate courses and to appoint some permanent staff, thus fulfilling the recommendation of a Royal Society report which nearly two years ago called for twenty more university posts in

biotechnology.

The research councils welcome the new grants, seeing no conflict between the grants committee's assessment of priority and their own. The Science and Engineering Research Council, in particular, welcomes the grants as a way of supporting staff and equipment which could not be paid for out of its research awards.

Judy Redfearn

Development and drugs

More not less

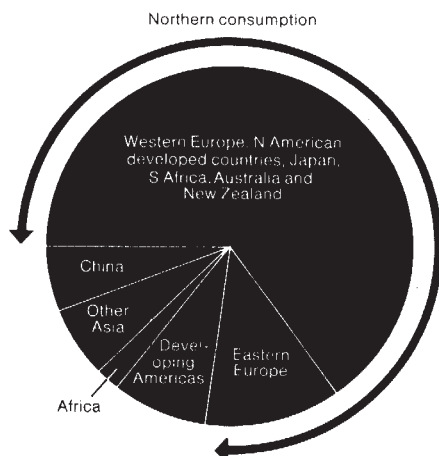
The latest shot in the long-running battle between the pharmaceuticals industry and its detractors, in which the health problems of developing countries provide the battleground, has been fired by the Office of Health Economics (OHE). Despite its governmental sounding title and Whitehall address, the office is sponsored by the UK pharmaceuticals industry and its main task is to carry out research on the economic aspects of medical care. Its latest contribution, *Medicines, Health and the Poor World* by David Taylor, is a response to recent criticisms of the industry's marketing practices.

The large multinational companies have been accused of over-aggressively selling unsuitable drugs in the developing countries, leading to only a minimal improvement in the health of the population and sometimes proving positively harmful. Chief among the industry's critics have been aid organizations such as Oxfam and War on Want and the pressure group Social Audit.

The report acknowledges that some drugs have been "inappropriately" sold in the past but claims that the industry itself is now more capable of policing its methods of promotion and that the important role of drugs in improving health care in developing nations may be obscured by concentration on abuse in some areas.

Although a typical poor nation may spend around a quarter of its central government health budget on pharmaceuticals, the report says, 60–70 per cent of the people do not have regular access even to the most basic drugs. So while it is important that those drugs now being sold to the "wealthier" members of developing societies are properly advertised and correctly used, it is even more important to find ways of getting the basic drugs to the mass of the population deprived of them. Whether the past performance of the multinational companies has contributed to the weaknesses of governmental health services in the developing countries, or whether the unavoidable difficulties have limited the ability of the drug companies to act effectively, remains a point of conflict.

Better distribution of a limited range of medicines and vaccines, together with research aimed specifically at new pharmaceuticals for the developing world, are the urgent needs, says OHE. The World



World pharmaceutical consumption for 1980 (total \$76,000 million, manufacturer's prices).

Health Organization (WHO) has put considerable effort into work on new medicines and vaccines for diseases such as malaria, onchocerciasis, leprosy and leishmaniasis, but the report argues that the commercial companies cannot take on work on new drugs for these diseases because of the harsh economic realities of the marketplace — the return from sales of "Third World-oriented" drugs would not cover the cost of their development.

Thus OHE argues that the major drug companies should be used as contractors by WHO and governments to carry out the expensive first stages of developing drugs for diseases in developing countries. It is argued that this approach would be the most cost-effective way for the developed nations to help the poorer nations achieve better health standards.

To put the expenditure involved in finding new drugs into context, the total research and development spending by the UK pharmaceuticals industry is around £300 million per year, compared with the £1,000 million spent annually by the UK government on overseas aid. So an increase of just 5 per cent in this aid budget would provide £50 million, which if spent on contracted research could support a significant effort to find drugs important to developing countries.

On the supply of existing medicines to the developing countries, the OHE report is critical of the WHO programme on "essential" drugs. As for policing the marketing and advertising methods used to promote drugs in the Third World, OHE sees the international code of marketing practice drawn up recently by the International Federation of Pharmaceutical Manufacturing Associations as the best hope for the future. The federation, it is argued, would be well able to regulate the activities of the major international companies because it is in the companies' interest to be seen to be acting responsibly throughout the world. WHO is seen by the industry as being open to political pressures and therefore ineffectual as a controlling body.

Charles Wenz

US weather and ocean research

Outlook bleak

Washington

The major controversy in Washington this week has been the fate of the federal budget, as congressional leaders and the President joust with each other, leaving the outcome in doubt. No less concerned are the atmospheric scientists and oceanographers, especially those involved in international programmes, many of which have been cut back in the President's proposed budget. Ocean and atmospheric research has always had friends in Congress, and they are trying to have the funds restored. But whether the money can be put back, and then retained in a final budget package, is in doubt.

Marked for the axe are funds for joint US-Canadian efforts to clean up the Great Lakes, half of the satellite capability that provides weather data to nations in Asia, Africa and South America, funds to study and prevent ocean dumping (including radioactive waste), and the World Climate Program, the successor to the Global Atmospheric Research Program, which is a major international effort run from Geneva.

The Canadian government has been very upset about the proposed cut-back in US efforts to help study, monitor and clean up the Great Lakes, to which the US government is committed under several agreements with Canada. Under the proposed Reagan budget, two laboratories would be closed down and one programme office severely curtailed.

One laboratory is run by the Environmental Protection Agency (EPA) at Grosse Ile, Michigan, near Detroit. The other is run by the National Oceanic and Atmospheric Administration (NOAA) at Ann Arbor, Michigan.

The Administration justifies the cuts by claiming that most of the research needed to identify Great Lakes pollution has been done, and that cleaning up is a responsibility of the states that border them.

But the Canadian government's view is that still more pollution problems are being unearthed, such as the identification and reduction of such toxic substances as mirex and dioxin which have recently been found in "hot spots" in all the lakes except Lake Superior, and that the international commitments are federal, not state, issues. According to one Canadian official the cuts amount to "an attempt by the United States to renege on its water quality agreement of 1978" with Canada. This is strong language, given the historical close cooperation between the two countries — at least until the advent of the acid rain dispute and the Great Lakes budget cuts.

The Administration also wants to cut NOAA funds for ocean dumping research and marine pollution generally. NOAA runs most of the government's research in

this area. The work concentrates on the area off the north-east coast of the United States which has the worst pollution problem, a major fishing ground and possible oil development. The rationale for the cut is that northeastern states should individually bear these costs. Congressmen seeking to restore the funds argue that these are national problems, especially in the light of renewed talk of disposing of radioactive and other forms of waste in the oceans.

For the second year in a row, the Administration has tried virtually to eliminate one of the ocean research programmes most popular on university campuses in the United States, the \$35 million per year Sea Grant programme. For 1983, the Administration proposes only \$1.7 million. However, as in 1982, representatives and senators are expected to get the funds restored. About half of the Sea Grant funds go for research, and the other half for community services related to the oceans issue.

The proposed Administration budget would also cut \$6 million of the \$7 million going to weather modification research, as part of a 40 per cent cut in atmospheric research funds. Likewise, a programme to upgrade the old weather radars on which continental US weather forecasting relies heavily is being slowed drastically. The Administration contends that weather is a local issue. Friends of NOAA counter, however, that weather modification and atmospheric research are important basic research programmes.

Of worldwide interest is a cut of \$24 million that would decrease the launch rate of the NOAA series of polar-orbiting weather satellites, so that there would be one instead of two in orbit at any given time. These satellites complement the existing two GEOS geosynchronous satellites that provide "side views" of the Earth's weather, including the familiar television-screen images, to many countries that have receivers.

Representative James H. Scheuer, chairman of the House Science and Technology Committee's subcommittee on natural resources, agriculture research and environment, is trying to get the second NOAA satellite restored.

Scheuer argues that while the geosynchronous satellites indicate current weather, mostly in the middle and lower latitudes, the polar-orbiting satellites are essential to forecasting worldwide. Only they can acquire the quantitative data needed for modelling the Earth's weather, and only they can track the Arctic and Antarctic air masses moving towards the inhabited regions. The NOAA series satellites are thus crucial to weather prediction in parts of the world, such as Asia, Africa and Australia, that other satellites do not "look at" as often as at North America and Europe.

Two satellites provide twice as many passes over a region as one, and so improve

daily forecasting. Two satellites also provide for redundancy. Scheuer says the Administration's cut will reduce the timeliness of forecasting and leave those parts of the world that are "neglected" by the rest of the system without any weather information at all when the remaining satellite breaks down — which happens often enough. Sometimes it takes a year to put a replacement for a broken satellite in orbit.

Finally, funds for the US contribution to the World Climate Program, a major international climatological effort run in cooperation with the World Meteorological Organization, have been cut from \$1.8 million to \$0.5 million.

Deborah Shapley

Lead in petrol

Clear risks

The British anti-lead lobby is flying high. Results now emerging from recent studies seem to be persuading scientific and medical opinion that the government should take further action on the lead content of petrol. Total elimination was advocated by many of the participants at an international conference organized last week by the Campaign for Lead Free Air (CLEAR).

Local councillors, civil servants and scientists from several countries heard evidence that children's IQ is impaired by low blood lead levels and that a significant contributor to lead burden is the lead in petrol. Des Wilson, chairman of CLEAR, is confident that the conference was given conclusive evidence that the position of the government is becoming increasingly isolated and untenable. The Lawther report, he said, had been completely discredited.

In 1980 a working party under the chairmanship of Professor P.J. Lawther advised that the evidence on danger from low blood lead levels was inconclusive. The Lawther working party also recommended that most effort be directed to reducing the lead in food and water, on the grounds that lead in petrol did not make the most significant contribution to the body lead burden.

The fear that blood lead concentrations of $300 \mu\text{g l}^{-1}$ and below may have neurological effects represents a gradual shift of opinion over the past two years. The official government position is that harmful effects do not occur below $350 \mu\text{g l}^{-1}$. The change of heart by the government which resulted in the decision last year to cut the legal amount of lead in petrol from 0.4 mg per litre to 0.15 mg per litre was in keeping with the Lawther case for a progressive reduction. Professor Lawther now says that the elimination of lead in petrol did not conflict with the findings of his working party and that "if there were damn all, nobody would be happier than me".

While there is no dispute about the toxic

effects of lead at high levels, controversy persists about the level at which harmful effects occur. After the Lawther working party reported, two members, Dr Richard Lansdown and Dr William Yule of the Institute of Psychiatry and Great Ormond Street Hospital for Sick Children, produced results suggesting a link between blood lead levels and IQ performance in children of the London suburb of Greenwich and claimed a significant difference of IQ performance with blood lead above and below $120 \mu\text{g l}^{-1}$ (*Devl Med. Child Neurol.* 23, 567-576; 1981).

At last week's symposium, Yule and Lansdown gave new evidence (as yet unpublished) that children's behaviour is related to blood lead levels — 19 per cent of those with blood lead levels above the average ($120 \mu\text{g l}^{-1}$) were overactive compared with only 4.9 per cent of those with blood lead levels below the average. Bad conduct, nervous tension and lack of concentration increased with higher blood lead levels. Professor Herbert Needleman (Children's Hospital, Pittsburgh) also reported an association between classroom behaviour, IQ performance and lead burden.

Others say, however, that there should be extreme caution in moving from an observation of correlation between childhood lead exposure and impairment of intellectual development to postulating lead as the cause of extreme behavioural and psychological damage. The social environment is an extremely important factor in IQ performance.

The conference was also told of the significant contribution of lead in petrol to lead burden, making the Lawther estimate of 10 per cent seem too low. One recently completed study by the EEC Joint Research Centre at Ispra, Italy, suggests that as much as 30 per cent of the lead in blood is derived from petrol. The study determined the contribution of lead from petrol by relying on the known abnormal isotopic composition of lead in the petrol in one region of Italy. The conference also heard that in the United States in the four years since the phasing out of lead in petrol (1976-80), blood lead levels had fallen by 36.7 per cent. Dr Clair Patterson (California Institute of Technology) said that "exhausts from leaded gasolines are the most serious sources of lead in people".

While opinion is still divided and the need for further research acknowledged, Professor Michael Rutter (Institute of Psychiatry, London), a member of the Lawther committee, came down firmly in favour of the elimination of lead from petrol: "the level of probability is such that I think it is worth acting on".

CLEAR moves confidently into battle having already captured the support of the Labour party. The National Executive Committee pledged last month that a future Labour government would eliminate lead in petrol. CLEAR's sights are now set on the party conferences where

it is confident of rallying support from the other opposition parties.

While much of the new evidence is not yet published, the Royal Commission on Environmental Pollution has now also launched an investigation into lead. It has heard evidence from the British Medical Association (BMA) supporting the link between low blood lead levels and impaired mental function. The BMA also appears to be convinced by the Ispra study.

The Royal Commission plans to investigate the sources of lead in the environment and its pathways to man. The commission also plans to "get to the bottom" of arguments about the technical and economic implications of reducing lead in petrol below that promised by the government last year. The commission may be confronted with strongly held opinions — Des Wilson described Associated Octel, the British company owned by Shell, BP, Chevron, Mobil and Texaco, which supplies the lead for petrol, as "the biggest mass child poisoners in the world today".

The oil companies say, however, that they will cooperate with the government if it calls for a ban on lead in petrol. The buck has now been passed.

Jane Wynn

Enter Exxon

New York

Cold Spring Harbor Laboratory on Long Island, New York, has agreed to a five-year "cooperative research agreement" with Exxon Research and Engineering Company. However, the financial terms of the deal are not being disclosed by Exxon, which is making its debut into the overcrowded world of biotechnology.

Under the terms of the agreement, up to six Exxon scientists will be assimilated into the Cold Spring Harbor staff to work full time on mutually agreed projects. In return, Cold Spring Harbor will select six postdoctoral fellows to participate in Exxon-funded research. In addition, Exxon's biosciences laboratory in New Jersey (which has never worked on molecular biology) may consult Cold Spring Harbor on various matters.

Exxon plans that its lawyers will visit the Cold Spring Harbor Laboratory regularly to keep abreast of latest developments. It will have exclusive rights to all patents derived from the research it funds, while Cold Spring Harbor Laboratory will retain rights to patents derived from work done by staff members not associated with Exxon projects.

Exxon will not ask those working on its Cold Spring Harbor Laboratory projects to defer publication of patentable inventions, as some industrial sponsors have requested of academic institutions relying on their funds.

Michael D. Stein

CORRESPONDENCE

Scientists who act

SIR — Sir Peter Medawar's review of Martin Gardner's book "Science: Good, Bad and Bogus" (*Nature* 28 January, p.351), criticizes people working in fringe science, scientists who step aside from their specialisms to apply their experience in other areas and the inexperienced whose opinions force them into a corner where they grimly hold on to their untenable theories. He paints a broad canvas of condemnation and thereby implies that most speculation is put forward by rogues motivated by profit or starved of recognition and that we on the receiving end are unable to distinguish between the possible, probable and downright foolish.

Sir Peter's attack questions the integrity of scientists whose imaginative speculations, right or wrong, stimulate action that eventually leads to the truth. He makes use of the reference to unnamed people — the astronomer, microbiologists — without stating his objections. Above all, Sir Peter forgets that absence of proof is not proof of absence. Truth prevails eventually and the speculations and experiments that lead to it are winnowed by time.

Sir Peter's preoccupation with the intellectual underworld does disservice to those who think as well as labour in their laboratories. Also to those who in these difficult times are without laboratories and who have only their thoughts to offer to the service of human understanding.

D.G. APPLIN

Leystonstone,
London, UK

Doves in false garb

SIR — The leading article "Doves in false garb" (*Nature* 18 February, p.542) asks "What other justification is there for the Medical Campaign against Nuclear War except that physicians prefer hobnobbing with other physicians than with the hoi polloi?" This is a cheap shot unworthy of *Nature*. Let us hear some specific criticism of the physicians' proposals, let us hear some refutation of the plentiful justifications offered for the campaign by the physicians, but let us not hear "Nyah, nyah, nyah."

As the present American government considers a nuclear war ever more wageable and winnable, physicians must also begin ever more often to picture themselves crawling about on radioactive heaps attempting to dispense medicine in the aftermath of victory. The hope for curing anybody is less than small. Physicians claim that a nuclear war is a terminal disease for entire nations. When a doctor helps us to not catch a disease this is called preventative medicine. Such help is not only legitimate for a doctor to offer but may be felt required by the Hippocratic oath. Physicians involved in the aforementioned campaign are doing what they perceive is their duty to keep their patients alive. Does *Nature* doubt this prognosis? Do you share the opinion of the Federal Emergency Management Agency to the effect that society (at least that of the United States) will totally

recover from a full-scale nuclear war within two to four years?

Nature concludes this commentary with the suggestion that if physicians, or any other professional group's members, seek to influence policy they should become members of parliament. This suggestion is nonsense. It is precisely because physicians are not politicians that they may help bring significant arms control into being. In past comment (*Nature* 294, 197; 295, 270) *Nature* has both stated the evident desirability of serious arms control negotiations and depicted the bumbling attempts of (US) politicians to initiate them. Atomic and hydrogen bombs now proliferate faster than ever before and members of parliament *et al.* have been as yet unable to inhibit this malignant growth. To stop and reverse the arms race will require the efforts of people in all walks of life acting in ways often without precedent, often compelled by personal, moral feeling.

Nature owes the living minds of its readers mature and honest evaluations of proposals from any campaign against nuclear war.

TOBIAS ISAAC BASKIN

Stanford Mid-Peninsula Chapter,
Physicians for Social Responsibility,
Stanford, California, USA

Mutualistic lives

SIR — In his News and Views article, "Mutualism: New Ecological Theories", May¹ asserts that "... conspicuous mutualistic associations tend to be tropical ones". May I suggest that conspicuousness is in the eye of the beholder? All the examples he cites involve animals. From ericaceous heaths of arctic-alpine regions, through vast acreages of boreal forest and temperate grasslands to tropical rain forests, the structure of the plant communities is governed, to those with the eye for them, by the conspicuous associations between plants and their mutualistic mycorrhizal fungi (see, for example, refs 2–6).

Mutualistic associations between plants and nitrogen-fixing prokaryotes and lichens are also important in tropical and non-tropical, especially sub-arctic, zones^{7,8}. Harley^{9,10}, myself¹¹ and others have bemoaned the slow recognition by taxonomists and ecologists of the importance of mutualistic fungi and bacteria to plants in all terrestrial ecosystems. Let us hope that another of May's assertions¹ — that empirical and theoretical studies of mutualistic associations are likely to be one of the growth industries of the 1980s — will be world-wide and encompass interactions between all five kingdoms of organisms¹².

D.H. LEWIS

The University,
Sheffield, UK

1. May, R.M. *Nature* 296, 803 (1982).
2. Harley, J.L. *The Biology of Mycorrhiza*, 2nd edn (Hill, London, 1969).
3. Janos, D.P. *Biotropica* 12, 56 (1980).
4. Malloch, D.W., Pirozynski, K.A. & Raven, P.H. *Proc. natn. Acad. Sci. U.S.A.* 77, 2113 (1980).
5. Malajczuk, N. & Lamont, B.B. in *Ecosystems of the World, Vol. 9B Heathlands and Related Shrublands: Analytical Studies* (ed. Specht, R.L.) 165 (Amsterdam, Elsevier, 1981).
6. Read, D.J. *Can. J. Bot.* (in the press).

7. Sprent, J.I. *The Biology of Nitrogen-fixing Organisms* (McGraw-Hill, London, 1979).
8. Hale, M.E. *The Biology of Lichens*, 2nd edn (Arnold, London, 1974).
9. Harley, J.L. *New Phytol.* 67, 979 (1968).
10. Harley, J.L. in *Ecological Aspects of the Mineral Nutrition of Plants* (ed. Rorison, I.H.) 437 (Blackwell, Oxford, 1969).
11. Lewis, D.H. in *Taxonomy and Ecology* (ed. Heywood, V.H.) 151 (Academic, London, 1973).
12. Margulis, L. & Schwartz, K.V. *Five Kingdoms. An Illustrated Guide to the Phyla of Life on Earth* (Freeman, San Francisco, 1982).

Kin to whom?

SIR — Kin selection is not normally used by students of mammalian social evolution to explain the sociobiology of a particular class^{1,2}. Contrary to the assertions by Packer and Pusey (*Nature* 22 April, p.740), Boorman and Levitt³ (p.250) derived the conclusion that "attempts to apply $k>(1/r)$ directly to the analysis of primate, carnivore or other mammalian societies should be evaluated very cautiously" from their thorough review of the literature on mammals.

Students of vertebrate social systems have recognized the importance of the "multi-male" group, selection for outbreeding, and reciprocal associations between unrelated males for social evolution in the mammals, and the implications of these classical features for the genetic and phenotypic structures of mammalian, including human, populations have been discussed^{4,9}. It is the persistence of these features that leads some primatologists to conclude that macaques and baboons, rather than gorillas or chimpanzees, are the richest models extant of hominid evolution^{10,11}. Further, numerous "theoretical analyses" have pointed out that "inclusive-fitness maximizing" and kin selection are not equivalent and that under certain competitive regimes, kin may be "ego's" worst enemies^{12–14}. Packer and Pusey's appreciation that individual reproductive "games" may not often be explained by kin selection for the mammals follows from the literature on behaviour and social evolution in that class.

CLARA B. JONES

Museum of Comparative Zoology,
Harvard University, Cambridge,
Massachusetts, USA

1. McCracken, G. & Bradbury, J. *Science* 198, 303–306 (1977).
2. Schwartz, O.A. & Armitage, K.B. *Science* 207, 665–667 (1980).
3. Boorman, S.A. & Levitt, P.R. *The Genetics of Altruism* (Academic, New York, 1980).
4. Trivers, R.L. *Q. Rev. Biol.* 46, 35–57 (1971).
5. Eisenberg, J.F. *Handb. Zool.* 10, 1–97 (1966).
6. Eisenberg, J.F., Muckenhirn, N.A. & Rudran, R. *Science* 176, 863–874 (1972).
7. Alexander, R.D. *et al.* in *Evolutionary Biology and Human Social Behavior: An Anthropological Perspective* (eds Chagnon, N.A. & Irons, W.) 402–435 (Duxbury, North Scituate, Massachusetts, 1979).
8. Brown, J.L. *The Evolution of Behavior* (Norton, New York, 1975).
9. West-Eberhard, M.J. *Proc. Am. Phil. Soc.* 123, 222–234 (1979).
10. Washburn, S.L. & DeVore, I. *Scient. Am.* 614, 2–11 (1961).
11. Tiger, L. *Men in Groups* (Random House, New York, 1969).
12. Alexander, R.D. *A. Rev. Ecol. Syst.* 5, 325–383 (1974).
13. Kurland, J.A. *Ethol. Sociobiol.* 1, 255–274 (1980).
14. Waded, M.J. & Breden, F. *Behav. Ecol. Sociobiol.* 7, 167–172 (1980).

NEWS AND VIEWS

The garden path

from Beverly Griffin

PERHAPS *Pilgrim's Progress* should after long neglect once again take its place on the literary scene, as required reading for today's molecular and cell biologists, and immunologists, instructing us to tread warily into the unknown, encumbered as it is with the snares, crevices and hidden dangers afforded by our modern technology. Not that there is anything sinister in the technology itself. Rather, we understand it too little and often use it too glibly. With due apologies to a distinguished group of scientists, I wish to quote a sentence from one of their papers to illustrate my point: "These results unequivocally demonstrate that chicken DNA contains nucleotide sequences related to (the) transforming gene of Y73 and that these endogenous DNA sequences . . . are different from the endogenous *src* related to *src* of RSV and do not correspond to any endogenous viral sequences"¹. (Both Y73 and RSV, Rous sarcoma virus, are RNA viruses capable of inducing fibrosarcomas in chickens, and their transforming genes are designated *yes* and *src* respectively.)

The data that indicated *yes* and *src* as being structurally unrelated were based on nucleic acid hybridization analyses in conditions that would generally be described as 'medium stringency'. They were said to "unequivocally demonstrate" and were no doubt correct, but nonetheless led to incorrect conclusions, as now shown in an article in this issue of *Nature* (p.205) by N. Kitamura and some of the authors of the above mentioned paper.

Kitamura *et al.* have sequenced the *yes* gene of Y73 and compared it with a (revised) sequence of *src*. Overall, the homology is about 70 per cent; but they find that, at the nucleotide level, the longest homologous stretch is only 17 residues, which explains why in the conditions used¹, no hybridization was previously observed. A great many of the non-homologies, however, prove to be third base changes, so the encoded protein structures are remarkably similar, allowing the authors to suggest that these two transforming viral genes were probably

derived from a common prototype sequence. The two protein sequences, although differing at the N- and C-termini, are almost indistinguishable over about 80 per cent of their length. Notably retained is a polypeptide stretch of over fifty amino acids containing the phosphorylated tyrosine residue that now appears to be a common feature of transforming genes of both RNA and DNA viruses.

In this context, another interesting paper by Stoner and Schlieff² questions whether the amino acid, but not the nucleotide, sequences of the *araC* genes in two strains of *Escherichia coli* have been conserved. Again, the data point to third base changes which would prevent hybridization but retain protein structure. Returning to the viral model, one can ask whether all the transforming genes of the class of viruses called retroviruses, when subjected to the same degree of scrutiny, might not prove to be similarly related. Take the *fps* gene in Fujinami virus (FSV), for example. Hanafusa *et al.*³ state that "the results described convincingly exclude the presence of RSV-related *src* gene in the genome of FSV. The lack of *src* was shown by oligonucleotide fingerprinting, molecular hybridization and immunoprecipitation with antiserum to the *src* gene product". The question now arises, will these techniques survive the nucleotide sequence test, or have they, in this instance, proved false in the hands of undeniably eminent scientists? At this stage, the possible relationship between viral transforming genes and oncogenes^{4,5} or that between the transforming genes of RNA and DNA tumour viruses seems a matter of conjecture. The question has to be raised of how much confidence one can put in negative data. With all the enthusiastic computing going on at many terminals, can a program be devised that will allow the accuracy of sequence rela-

tionships to be estimated, particularly with respect to hybridization?

Before this becomes possible, one presumably needs to understand what hybridization measures. Is it, for example, 90 per cent hydrogen bonding (G-C carrying more weight than A-T) and 10 per cent pi-electron interaction (purine-purine interactions more significant than pyrimidine-pyrimidine)? What effect does the location of imperfections in homologies have on the hybridization? How effectively does intramolecular hybridization compete with intermolecular hybridization and under what conditions? Since much weight is put on the data derived from hybridization between partly complementary sequences, shouldn't we know the answer to questions such as these? The future clearly will involve searching for specific cellular genes with the aid of this very powerful, but also less than well understood, technique.

The literature contains a smattering of good papers on the subject. To quote but a few that help to clarify matters, a start can be made with a paper by Howley *et al.*⁶. The authors chose three related and well defined (that is, sequenced) DNA viruses of different host origins to compare by hybridization — SV40 from monkey, BK virus from man and polyoma virus from mouse cells. Under stringent hybridization conditions, very little homology was observed, and this was assigned to the region encoding the viral capsid proteins, not a region of primary importance in the pathology of these viruses. By sequentially decreasing the stringency, however, homologies began to appear elsewhere, such that the coding regions of SV40 and BKV could be observed to be remarkably similar, and also similar to about 75 per cent of polyoma virus. The validity of these hybridization studies was confirmed by a direct comparison of their DNA sequences⁷, and evolutionary relationships among the viruses were proposed^{6,8}. What was the outcome of this less qualitative and rather more detailed approach to hybridization than is usually seen in the literature?

Beverly Griffin is in the Imperial Cancer Research Fund Laboratories, Lincoln's Inn Fields, London WC2A 3PX.

One outcome was to suggest that immunological relationships should exist among some of the viral proteins other than the structural proteins. McCormick *et al.*⁹ thus looked for, and (in contrast to previous negative reports) found, immunological cross-reaction between the large T(tumour) antigens of SV40 and polyoma virus. They conclude their paper with the warning note, "that a cross-reaction between two proteins of related function and extensively homologous primary sequences should prove so elusive sounds a note of caution in the interpretation of negative results when immunological techniques are used to search for homologies between polypeptides".

At a simpler level, but perhaps the one from which quantitative understanding of hybridization will ultimately come, Gillam *et al.*¹⁰ have studied melting temperatures between duplexes of oligo(T) and oligomers such as A₂TA and A₂GA. A difference of several degrees in the relative stabilities of the duplexes was found, with the rather surprising result that the former mismatched double helix appeared to be more stable than the latter one. As the length of the oligomers was increased from the septamer to the nonamer, this difference diminished. (Among many interesting conclusions, they note that an oligodeoxyribonucleotide is less likely to form a mismatched structure with DNA than with RNA.) When homologies between dissimilar but related species are compared, is not hybridization essentially a result of a composite of melting out and reforming short duplexes?

A related approach has been taken in a series of stimulating papers by Wallace *et al.*¹¹, who investigated hybridization by using oligodeoxynucleotides as 'hybridization probes' for specific genes. By modification of the conditions under which hybridization was carried out, they could discriminate between the formation of perfectly matched and single-mismatched duplexes involving rabbit β -globin DNA and fourteen-long oligomers. To end on another 'perhaps', studies such as the elegant ones mentioned above may ultimately lead to a series of rules for the selection of the appropriate hybridization conditions that will allow interactions, whether between nucleic acids, or proteins, or even between the two, to be predicted accurately, even before detailed sequences become known. □

Fission, fusion and fracture

from Robert W. Cahn

CONFIDENCE in the intrinsic feasibility of an economically effective fusion reactor is now sufficiently widespread among those responsible for research policy that technological aspects of fusion reactor design are receiving detailed attention. One straw in the wind is the recent publication of the Proceedings of the Second Topical Meeting on Fusion Reactor Materials¹.

While about a quarter of the papers deal with such matters as tritium and deuterium recycling, breeding blankets and their constituent materials (especially lithium), and alloys for magnet windings, the central concern is the 'first wall' — the inner surface of the torus which is exposed to the full intensity of the 14 MeV neutrons resulting from the D-T fusion reaction. Schiller² surveys the various dangers that beset a first wall: swelling and embrittlement due mainly to helium generated *in situ* by a nuclear reaction due to neutrons, especially with nickel; erosion by sputtering due to neutral atoms escaping from the plasma, which might remove as much as 1 cm from the surface during a 15-year reactor lifetime; and thermal fatigue because the operation of the reactor will be pulsed.

A number of alloys have been proposed for the first wall, including some exotic ones. Thus several investigators have seriously proposed the use of metallic glasses which typically crystallize at 350–600°C: Emmoth *et al.*³ have shown that the familiar glass Fe₄₀Ni₄₀B₂₀ sputters three times more slowly than Fe-Ni-Cr (crystalline) stainless steel. (This is interesting from a fundamental point of view also since no rigorous comparative measurements have yet been made of sputtering yields from the same alloy in glassy and crystalline forms.)

However, the 'prime candidate materials' for the first wall appear to be an Al-Mg-Si alloy and variants of type 316 stainless steel (with ~15 wt per cent each of Ni and Cr). A great deal of research on 316 steel was reported at Seattle: it is a familiar material, used for components of fast fission reactors as well. The most impressive of the papers on 316 steel variants are a group devoted to steels incorporating a fraction of a per cent of titanium carbide. One paper⁴ from the very active research group at Jülich in West Germany compares the behaviour of 316 steel with and without TiC. The unirradiated creep rupture life at constant stress at 700°C and the reduction of creep rupture life due to helium ion implantation (which simulates what happens in a fusion reactor first wall) were very much better in the modified steel. This is plainly due to a fine distribution of TiC in grain interiors and the ability of these TiC particles to trap helium bubbles away from the grain

boundaries where alone they produce creep embrittlement. This is an example of the increasingly common metallurgical ploy of setting one impurity trap to combat the effects of another.

A steel almost identical to that studied at Jülich is the American 'prime candidate path A alloy'. Arnberg *et al.*⁵, of MIT, show that if this alloy is made by rapid solidification processing (flakes frozen at ~10⁵ K s⁻¹ and then hot extruded), a fine grain size (~10 μ m) and extremely fine TiC particles result. This material, when helium-bombarded, proved to be very resistant to swelling at ~600°C, even more so than the same alloy made by conventional methods. The TiC, according to the investigators' electron micrographs, nucleates numerous helium bubbles (especially the more numerous TiC particles in the rapidly frozen alloy) and these bubbles all remain very small, presumably under high pressure. Megusar *et al.*⁶ (also at MIT) have prepared versions of this steel with triple the TiC concentration (only rapid freezing will give a uniform distribution of particles at this concentration) and this is expected to be even more resistant to swelling. The corresponding embrittlement tests have not yet been undertaken.

The 316 steels are austenitic (face-centre cubic in crystal structure). Ferritic (body-centred cubic) Fe-Cr steels without nickel have also been examined. They have the advantage of generating less helium from (n, α) reactions than do 316 steels. These steels are hardened but also quite severely embrittled by irradiation⁷, but Japanese work showed that a high Cr content (15 wt per cent) minimizes embrittlement⁸.

An interesting German paper⁹ shows that ordinary pressure vessel steel (without Ni or Cr) embrittles much less when irradiated at the service temperature of 290°C (in a PWR reactor pressure vessel) than when irradiated at ambient temperature, in the sense that the ductile-brittle transition temperature is less affected. However, if the steel contains a small amount of copper (a notorious source of contamination from recycled scrap steel), then embrittlement at service temperature is much more substantial. Here, an involuntary trace element, unlike the intentionally added Ti in 316 steel, has dangerous effects; but it is not difficult to control copper levels in high-grade steel.

Pressure vessel steels are currently being analysed in great detail in terms of linear elastic fracture mechanics, and the characterization of such steels goes a long way beyond a study of the effect of

Robert W. Cahn is Professor of Metallurgy at the University of Paris-Sud, Orsay 91405.

1. Yoshida, M., Kawai, S. & Toyoshima, K. *Nature* **287**, 653 (1980).
2. Stoner, C.M. & Schlieff, R. *J. molec. Biol.* **154**, 649 (1982).
3. Hanafusa, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3009 (1980).
4. Weinberg, R.A. *Trends biochem. Sci.* **135**, April (1982).
5. Bishop, J.M. *Scient. Am.* **69**, March (1982).
6. Howley, P.M. *et al. J. biol. Chem.* **254**, 4876 (1979).
7. Soeda, E. *et al. Nature* **283**, 445 (1980).
8. Soeda, E. *et al. Nature* **285**, 165 (1980).
9. McCormick, F., Lane, D.P. & Dilworth, S.M. *Virology* **116**, 382 (1982).
10. Gillam, S., Waterman, K. & Smith, M. *Nucleic Acids Res.* **2**, 625 (1975).
11. Wallace, R.B. *et al. Nucleic Acids Res.* **9**, 879 (1981), and references therein.

irradiation on ductile–brittle transition temperatures. It is possible, with a very high degree of confidence, to calculate for any part of a pressure vessel to what size a crack must grow before it can present any danger. The basis for these calculations have been outlined in technical terms by Nichols and Cowan¹⁰ and, with his incomparable clarity of exposition, by Cottrell in his popular book on reactor safety¹¹. While a failure of a fusion reactor first wall would certainly represent a far less serious accident than the fracture of a PWR pressure vessel (which, of course, has never happened), no doubt fracture mechanics will in due course be applied to the fusion situation as well, since it

represents the guarantor of nuclear safety where the prevention of fracture is concerned. □

1. Proc. Second Topical Meeting on Fusion Reactor Materials, Seattle, August 1981 *J. nucl. Mater.* **103 & 104** (1981–2).
2. Schiller, P. *J. nucl. Mater.* **103/4**, 75 (1981).
3. Emmoth, B. *et al. J. nucl. Mater.* **103/4**, 393 (1981).
4. Kesternich, W. & Rothaut, J. *J. nucl. Mater.* **103/4**, 845 (1981).
5. Arnberg, L., Vander Sande, J.B., Frost, H.J. & Harling, O.L. *J. nucl. Mater.* **103/4**, 1069 (1981).
6. Megusar, J., Arnberg, L., Vander Sande, J.B. & Grant, N.J. *J. nucl. Mater.* **103/4**, 1103 (1981).
7. Kibueh, R.L., Vitek, J.M. & Grossbeck, M.L. *J. nucl. Mater.* **103/4**, 887 (1981).
8. Suganuma, K., Kayano, H. & Yajima, S. *J. nucl. Mater.* **105**, 23 (1982).
9. Ahlf, J. & Schmitt, F.J. *J. nucl. Mater.* **105**, 48 (1982).
10. Nichols, R.W. & Cowan, A. *Phil. Trans. R.Soc. A299*, 227 (1981).
11. Cottrell, A. *How Safe is Nuclear Energy?* 60 (Heinemann, London, 1981); see also *Nature* **293**, 763 (1981).

Star formation in galactic spiral arms

from Gerard Gilmore

TWO-THIRDS of all bright galaxies show spiral structure, with the arms being made visible by young luminous stars and the giant gas clouds which are the sites of continuing star formation. The relation between the existence of the arms and the young stars which delineate them is analogous to the relation of the chicken to the egg. Do the galactic density waves which sweep material into the arms purely by chance create the ideal conditions for star formation, and thereby provide an optical tracer of the density enhancement, or is star formation a more fundamental requirement for the existence of the spiral arms? While current theories of galactic spiral structure suggest that the star formation is incidental, a recent numerical study of the effect of star formation on the appearance of a rotating galactic disk

suggests that, in some cases at least, bursts of star formation may in fact generate the observed structure (Seiden, Schulman & Feitzinger *Astrophys. J.* **253**, 91; 1982).

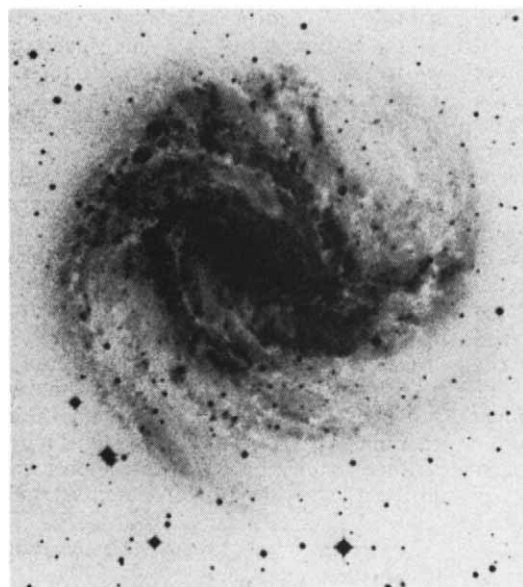
There are currently two widely discussed theories of galactic spiral structure. The first, originated by Lindblad in 1941 and developed extensively by C.C. Lin and associates, involves the response of a disk to an imposed spiral perturbation to the galactic gravitational field. As the stars and gas respond to this perturbation, they form a spiral density perturbation, which sustains the gravitational perturbation. The density wave is therefore self sustaining. The important difficulties with this theory are to find an origin for the perturbation, to maintain it against energy loss to resonances in the stellar orbits and to prevent its destruction by winding up. In general, the stars and gas outside the nuclear regions of galaxies have flat rotation curves, implying strongly

differential rotation with distance from the galactic centre. This in turn will smear out any material structure in a few rotation periods. Spiral galaxies are, however, typically 50 rotation periods old. So what maintains the spiral structure?

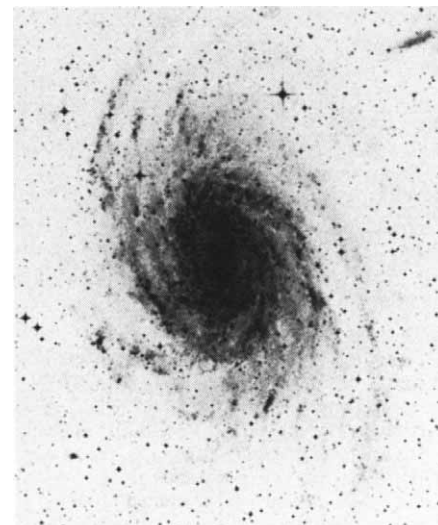
In the second theory, extensively developed by Alar Toomre and collaborators, the spiral density perturbation is generated as a consequence of dynamical interactions with nearby companion galaxies. His numerical simulations of such interactions show a strong density wave propagating through the disk galaxy as a consequence of internal dynamical resonances (called 'swing amplification' by Toomre) initiated by quite small density perturbations. This model can generate density waves, but requires them to be transient. The high proportion of disk galaxies showing spiral structure is then a consequence of the high rate of dynamical interactions between galaxies. The spiral structure of our Galaxy becomes a consequence of the orbital properties of the other galaxies in the local group.

While proponents of both these theories are able to generate models in striking agreement with observation, what of isolated galaxies, where no obvious origin for the spiral density perturbation is apparent? It is here that the model of Seiden *et al.* is of interest. They have extended earlier work by Seiden and Gerola (*Astrophys. J.* **233**, 56; 1979; *Nature, News and Views* **283**, 133; 1980) modelling the effect of star formation on the appearance of a galactic disk by including a simple approximation to current ideas of self-propagating star formation (SPSF). In this model, a few very massive stars formed in a dense molecular cloud evolve rapidly, and generate strong shock waves from stellar winds or supernovae explosions. As these shocks propagate through the remaining dense cloud, they initiate further compression with consequent enhanced star formation, leading to more massive

Gerard Gilmore is at the Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ.



Left, M83, a large, almost face-on spiral galaxy perhaps 10 million light-years away. The complex spiral arms are patterned with masses of dust, some of which seem to occur in giant shells many hundreds of light years across, and contain large numbers of young, luminous stars and masses of hot, glowing gas. At the centre of the galaxy is a source of radio waves. Right, one of the largest barred spiral galaxies, of visual magnitude 10.6 and an angular diameter of 9 arc min, in the constellation of Pavo.



Photography by Photolabs, Royal Observatory, Edinburgh

stars, and so on (Lada, Blitz & Elmegreen in *Protostars and Planets* ed. Gehrels, 341; 1978). The positive feedback rapidly destroys the conditions for further star formation, leaving a proportion of the cloud as gas. Eventually, this will cool and recollapse, ready for another burst of star formation.

Seiden *et al.* model this SPSF process by allowing random star formation to initiate bursts of subsequent formation, and explicitly including a cooling time for the remaining gas. As a burst of star formation propagates outwards, differential galactic rotation smears the expanding shell into an arc. The arc will then grow until it meets a region in which star formation occurred less than the recovery time before. With this simple view, the simulations are able to generate two-armed spiral galaxies allowing the initial star formation to occur randomly (hence their name: stochastic self-propagating star formation — SSPSF), and without having to impose an initial spiral density perturbation. They note that their process may in fact be an appropriate mechanism to stabilize a spiral density wave for many galactic rotation periods.

Gerola *et al.* (*Astrophys. J.* **242**, 517; 1980) have also pointed out a further important property of models of this type for small galaxies. If the recovery time for the gas is comparable to the galactic rotation period, the star formation will occur in delayed but periodic bursts. Large bursts of star formation have in fact been invoked by many authors for several years to explain the properties of small irregular galaxies and anomalously blue galaxies. While most of these bursts can be explained, as with spiral arms, by

dynamical interactions of nearby companions, a few cannot. As Seiden *et al.* point out, it is not clear how to distinguish observationally between periodic bursts of star formation for the few isolated galaxies, as predicted by their model, and a non-periodic phenomenon.

Perhaps the best summary of the current state of spiral density wave theory is by Toomre (see *Structure and Evolution of Normal Galaxies* eds Fall & Bell, 111; 1981): "It seems clear now that the spiral structure of galaxies is a complex riddle, without any unique and tidy answer". □

The great Japanese IQ increase

from Alun M. Anderson

THE mean IQ of the Japanese has risen spectacularly this century and, among the younger generation, is now the highest in the world, some 11 points above the mean IQ of the United States and other Western populations, according to a report by Richard Lynn in this issue of *Nature* (p.222). Analysis of results from the American Wechsler Intelligence Scale for Children shows that more than three quarters of the Japanese younger generation have a higher IQ than the average American or European. Lynn points out that with this difference in mean IQ there will be a far greater proportion of the Japanese population with a high IQ — 10 per cent of the Japanese younger generation should have IQs above 130, the levels generally found among professional groups such as research scientists, engineers, lawyers and doctors, compared with only 2 per cent of the American population. If, indeed, any link between IQ and intellectual achievement is accepted then the future implication for societies which are becoming increasingly dependent upon technological innovation may be profound.

How can this remarkable increase be explained? The dramatic changes in Japanese society that have occurred in the last century — a mixing of previously isolated peasant communities in the massive post-war urbanization, rapid economic growth and accompanying improvements in welfare, health and education and exposure to Western culture and ways of thought — probably provide a large part of the explanation. By comparing the group of Japanese born in 1910–1945 (mean IQ 102–105, only a little above the American level) with that born between 1946 and 1969 (mean IQ 108–115) it is possible to find evidence of changes in that period in almost every factor, genetic, nutritional, educational and social, that has been supposed to affect IQ. Some of these can be documented by a few key measures of Japanese society.

Japan retained a feudal society much later than the West. Until 1868, when the Emperor was restored to power after nearly three centuries of rule by the military Shogunate, the great mass of the population were peasant farmers, tied to the land and marrying within the local

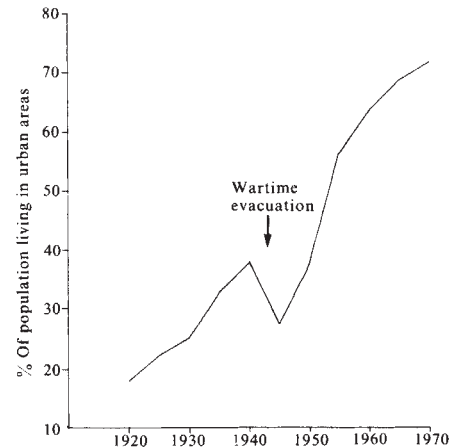
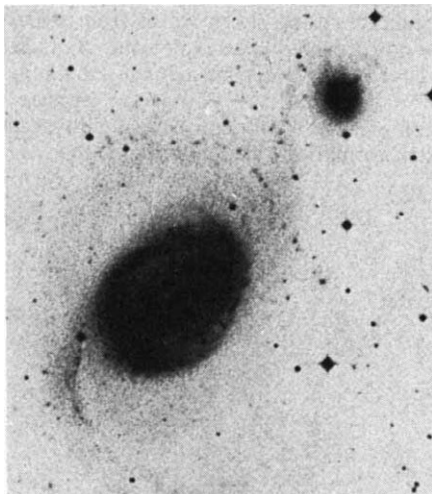


Fig.1 Increase in the urban population of Japan between 1920 and 1970 (from ref.1).

communities. In the late nineteenth and twentieth century a movement of the population towards the cities began, but it was not until the middle of this century that urbanization gained pace. In the period between 1930 and 1960 almost 40 per cent of the population moved from the country to the cities! (see Fig. 1). The remoter rural areas of Tohoku, Hokuriku, San'in, Shikoku and Kyushu were drained of their population, whole villages sometimes being abandoned, as Tokyo, Nagoya and Osaka emerged as massive industrial centres. In the new urban areas rural groups that had been isolated for many centuries now mixed and inter-marriage between individuals from different parts of the country became common. As a result the Japanese population has almost certainly become more outbred — a supposition supported by data on consanguineous marriages gathered in the exhaustive genetic studies of the populations of Hiroshima and Nagasaki carried out by the Atomic Bomb Casualty Commission². The change is one factor that could have contributed towards the IQ increase; that inbreeding can reduce cognitive performance, including IQ scores, is well documented (see ref. 3, for example).

Raised nutritional standards are very likely to have had major effect on IQ. Lynn

Alun M. Anderson coordinates the News and Views columns of *Nature*.



The larger of the two galaxies is NGC1512, consisting of a cigar-shaped bar of stars surrounded by a bright ring from which the spiral arms trail. The small elliptical companion, NGC1510, contains many young stars and clouds of hot gas, features which are very rare in galaxies of this type. It is thought to have 'rejuvenated' itself by scooping up gas from the cloud around its larger companion; the newly acquired material has since formed stars.

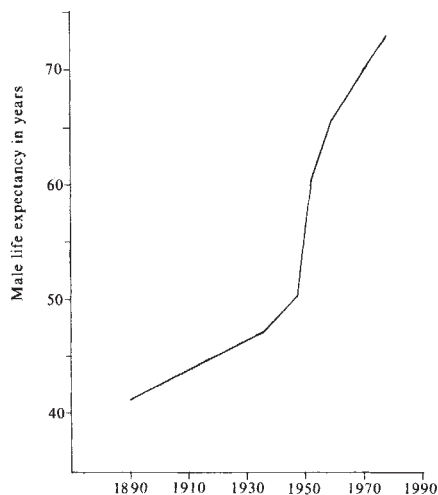


Fig.2 Male life expectancy in Japan between 1890 and 1977 (from ref.8).

points out their importance and documents them by noting the increase in birth weights. His conclusion is supported by the recent increases in the height of the Japanese and the rapid rise in life expectancy.

Japanese males born in 1960 are on average 5.2cm (2.1 inches) taller than those born in 1940. A similar increase was observed in Western Europe in the late 19th and early 20th century⁴ and was closely linked to improving economic conditions. Interestingly, the change is not simply one of height but of proportion, for leg length has increased by 4.3 cm while sitting height (from base of spine to top of head) has increased by only 0.9cm in the same period⁵. (Thus the popular, but erroneous, Japanese view that they have grown taller by adopting the habit of sitting on Western chairs rather than sitting with legs folded underneath the body on a *tatami* mat.)

The rapid rise in life expectancy (to 72.7 years for men and 78.9 years for women in 1977 compared to 46.9 years and 48.6 years in 1935, see fig. 2) has produced yet another first for the Japanese. In 1977 they became the world's longest lived race and their life

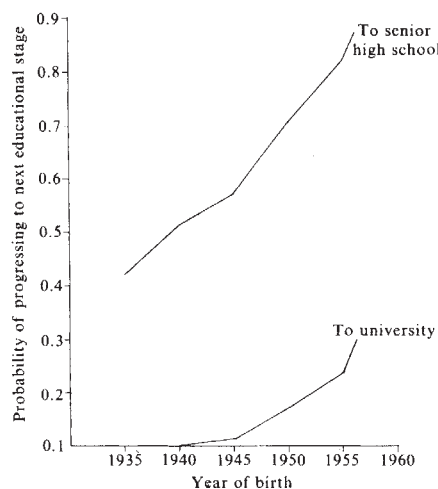


Fig.3 Fraction of middle school students going on to senior high school and to university or higher education college according to year of birth (from ref.9).

expectancy has continued to rise since then.

Of more obvious significance for the rise in IQ may be the very great increase in educational opportunities since the 1930s. School attendance is well known to have a substantial effect on IQ levels⁶. Among those born in 1948 and graduating from junior high school in 1962, only 51.5 per cent went on to senior high school and 10.3 per cent to college. By 1972, 87 per cent of junior high school students were progressing to senior high school and 30 per cent going on to college (see Fig. 3).

Nor, indeed, is this the first time that such a rise in IQ has been recorded. In the early days of American IQ testing many new immigrant groups were found to have low IQ (admittedly on much less sophisticated tests) but as they became accustomed to American life their IQs rose⁷. A similar effect may have occurred as the Japanese have become used to tests devised for a very different culture.

The change in IQ may, then, be explained in a multitude of ways. What remains a puzzle is why the rise should have

been to a level significantly higher than that found today in the United States. Once again, given the massive cultural differences between the West and Japan and the difficulties of devising any test that does not favour one culture or another there would seem to be endless opportunity for speculation. Whether the difference in IQ represents a real difference in 'intelligence' or simply implies that the Japanese are better at IQ tests, remains open to question. At present, there is little more to say than that. □

1. Bureau of Statistics, Office of the Prime Minister, 1970 Population census of Japan (1971).
2. Schull, W.J. & Neel, J.V. *The Effects of Inbreeding on Japanese Children* (Harper & Row, New York, 1965).
3. Vogel, F. & Motulsky, A.G. *Human Genetics*, 428 (Springer Berlin, 1979).
4. Vogel, F. & Motulsky, A.G. *Human Genetics* 185 (Springer, Berlin, 1979).
5. Ministry of Education Survey (1977).
6. Jensen, A.R. *Straight Talk about Mental Tests*, 183 (The Free Press, 1981).
7. Gould, S.J. *The Mismeasure of Man* (Norton, 1981).
8. White Paper on Health and Welfare, Ministry of Health and Welfare (1977).
9. White Paper on Education, Ministry of Education (1977).

The 'fold-back' elements of *Drosophila*

from D.J. Finnegan

GENOMIC rearrangements and DNA transposition have been observed in eukaryotes so often as to be a novelty no longer. Their significance and the various mechanisms by which they occur are far from clear, however, and at this stage all examples bear detailed investigation. The best characterized eukaryotic transposable elements are the 'copia-like' elements¹⁻³ of *Drosophila melanogaster*. They are not the only transposable elements in *D. melanogaster* and in this issue of *Nature* (p.201), Potter presents the complete DNA sequence of a member of a striking different class of transposable elements — the 'fold-back' elements.

When DNA from any eukaryote is denatured and allowed to re-anneal at low concentration the first sequences to become double-stranded are those that occur as inverted repeats. The repeats may be immediately adjacent or separated by up to several thousand base pairs and, as the reaction is intramolecular, the structures formed are called 'snap-back' or 'fold-back' structures. Schmidt *et al.*⁴ and Baker and Thomas⁵ have estimated that there are 2,000–4,000 pairs of inverted repeats, or potential fold-back structures, in *D. melanogaster* which comprise about three per cent of the total genome. To study the properties of the fold-back structures Potter *et al.*⁶ screened a library of cloned

fragments of *D. melanogaster* DNA with a probe enriched for the stems of fold-back structures. They obtained one clone containing an inverted repeat and used it to select a number of others containing related fold-back (FB) elements. The sequence of one of these, FB4, is reported in this issue.

Potter *et al.*⁶ also determined the chromosomal distribution of FB elements by hybridizing an FB probe to polytene chromosomes of larvae from different strains. About 30 hybridization sites were found in each case but their actual positions differed, indicating that FB sequences can move about the genome, that is, that they are transposable elements.

There are approximately 25 families of copia-like elements in the *D. melanogaster* genome. The members of each family are well conserved and are bounded by direct repeat sequences a few hundred nucleotides long, at the ends of which are short inverted repeats. The sequence of the direct repeats varies little within a family and, in all cases looked at so far, the repeats at the end of any one element are identical. In contrast, FB elements vary greatly in their inverted repeat sequences and the DNA separating them^{6,7}. The inverted repeats have a sequence organization similar to that of satellite DNA⁷ and those of FB4 are made up largely of tandemly repeated copies of a 155 base pair repeat unit which itself contains smaller sub-repeats⁸. Inverted repeats of different FB elements vary in length from a few hundred to a few

D.J. Finnegan is in the Department of Molecular Biology, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR.

thousand base pairs, presumably reflecting variation in the number of repeat units present⁷. The inverted repeats of FB4 differ in this way, with the right-hand repeat being longer than the left, and Potter points out that length variation of this type could be generated by mismatch recombination between repeats of different FB elements. In addition to these gross differences, the two repeats also differ by base substitutions and deletions/additions. The most highly conserved regions are at the extreme ends of FB elements. This is not surprising since these regions must presumably interact with any proteins involved in transposition.

The central region of most FB elements is made up of sequences related to the inverted repeats themselves⁷. FB4 is unusual in that its terminal repeats are separated by about 1,750 base pairs of unrelated DNA containing a potential gene coding for a 148 amino acid polypeptide. Most of this central region may itself be a transposable element since 33 nucleotides at its right-hand end occur as a near-perfect inverted repeat a short distance from the left-hand end (terminal inverted repeats are characteristic of all known prokaryotic and eukaryotic transposable elements).

The mechanism by which *copia*-like elements are able to redistribute themselves around the genome is by no means clear but the many properties they have in common with proviruses³ suggest that it could involve circular intermediates⁹ produced by a transcription-reverse transcription cycle. This is unlikely to be true of FB elements, however, and Potter (p.201) suggests that a transposase similar to those of bacterial transposons is involved and might be coded by the central region of FB4. Against this one has to remember that FB4 is the only FB element known to contain this sequence and its putative gene product might equally well be a transposase for the potential transposable element comprising this region.

The role (if any) of FB elements is unknown. Goldberg *et al.*¹⁰ reported that an FB element lies at one end of one of the very large transposable elements (TE elements) described by Ising⁸. These are

several hundred kilobases long and in some cases can be detected cytologically. The DNA involved can include respectable components of the genome, such as the white eye gene (*w*), unlike *copia*-like elements which seem to be genomic parasites. This raises the interesting possibility that any sequence bounded by two FB elements may potentially be transposable.

Inverted repeats can be detected in rapidly re-annealing DNA from many species and one may wonder whether these are similar to FB elements in the same way that endogenous proviruses seem to correspond to *copia*-like elements. In humans, the bulk of inverted repeat

structures are comprised of two copies of the 300 base pair 'Alu' sequence in opposite orientation and separated by variable lengths of unrelated DNA¹¹. Alu sequences make up a substantial component of human dispersed repetitive DNA and have counterparts in other species¹². Jagadeeswaren *et al.*¹³ have suggested that Alu sequences can themselves transpose but there is no indication that they behave like FB elements. It does not follow that humans and other vertebrates do not contain such elements, however, since even in *D. melanogaster* they make up less than two per cent of inverted repeats. All of this goes to show that there is still much work to be done in the study of repetitive DNA. □

Seamounts and flexure of the lithosphere

from A.B. Watts

MORE than 75 years have passed since the first gravity measurements showed that the Hawaiian islands, in the Pacific Ocean, were associated with large positive gravity anomalies of up to a few hundred mGal. The American geodesists, Hayford and Bowie¹, showed in 1912 that these anomalies could be substantially reduced by taking into account isostasy. The form of isostasy they preferred was the Pratt model, which considered that the islands were locally compensated by lateral variations in crustal density. The model was popular with geodesists since it enabled them to readjust the triangulation system of the US: a task of immense practical importance in mapping and surveying. It was of comparatively little interest, however, to a leading geologist at the time, Barrell, because it did not account for the lateral strength of the crust. Putnam pointed this out to Hayford and Bowie who subsequently amended the Pratt model by extending the compensation from beneath the islands into surrounding regions. They showed that a regional model of isostasy could reduce the gravity anomalies over the Hawaiian islands further than could the Pratt model.

But it was not until 1926, when Vening Meinesz² made the first pendulum gravity measurements in a submarine, that it was possible to prove regional, rather than local, isostasy for the Hawaiian islands. The measurements showed that the positive anomalies over the islands were flanked by negative anomalies of up to 100 mGal, which Vening Meinesz first interpreted as a result of downbending or flexure of the crust due to the load of the Hawaiian volcanoes. He used a model of an elastic plate overlying a weak fluid substratum, similar to one developed by Hertz³ to model the flexure of ice on ponds by skaters. Subsequently, Vening Meinesz demonstrated that a flexure model of

isostasy reduced the gravity anomalies over the islands further than could either the Pratt model, favoured by Hayford and Bowie, or the Airy model, preferred by some European geodesists.

Since the development of the concept of plate tectonics the flexure model of isostasy has assumed a special significance. Plate tectonics views the outer layer of the Earth, or lithosphere, as consisting of several large plates which converge, move apart and slide past each other. The plates therefore correspond to the rigid layer argued for earlier, on isostatic grounds, by Barrell and Vening Meinesz. An important assumption of applying plate tectonics to the geological past is that the plates behave rigidly for long periods of time. Thus, one of the objectives of current flexure studies has been to determine the thickness of the rigid layer and whether it varies on long geological time scales.

The numerous oceanic islands and seamounts that occur in the world's oceans have proved particularly satisfactory loads for flexure studies. They are mainly of volcanic origin, form relatively quickly on the plates (as evidenced by the short times between eruptions) and occur in a variety of tectonic settings. Since WW II there have been a number of geological, geophysical and geodetic investigations of seamounts and oceanic islands so there is now a large amount of bathymetry, gravity, geoid and recent crustal movement data for them. By comparing these data with predictions based on elastic plate models it has been possible to estimate the flexural rigidity, and the equivalent elastic thickness of oceanic lithosphere, in different tectonic settings in the oceans⁴⁻⁷.

Flexure studies at seamounts and

1. Rubin, G.M. *et al.* *Cold Spring Harb. Symp. quant. Biol.* **45**, 619 (1981).
2. Tchurikov, N.A. *et al.* *Cold Spring Harb. Symp. quant. Biol.* **45**, 655 (1981).
3. Finnegan, D.J. *Nature* **292**, 800 (1981).
4. Schmidt, C.W., Manning, J.E. & Davidson, N. *Cell* **5**, 159 (1975).
5. Baker, R.F. & Thomas, C.A. in *DNA Insertion Elements, Plasmids and Episomes* (eds Bukhari, A.I., Shapiro, J.A. & Adhya, S.L.) (Cold Spring Harbor Laboratory, New York, 1977).
6. Potter, S., Truett, M., Phillips, M. & Maher, A. *Cell* **20**, 639 (1980).
7. Truett, N.A., Jones, R.S. & Potter, S.S. *Cell* **23**, 753 (1981).
8. Ising, G. & Black, K. *Cold Spring Harb. Symp. quant. Biol.* **45**, 327 (1981).
9. Flavell, A.J. & Ish-Horowitz, D. *Nature* **292**, 591 (1981).
10. Goldberg, M.L., Paro, R. & Gehring, W.U. *EMBO J.* **1**, 93 (1982).
11. Deininger, P.L. & Schmid, C.W. *J. molec. Biol.* **106**, 773 (1976).
12. Jelinek, W.R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **77**, 1398 (1980).
13. Jagadeeswaren, P., Forget, B.G. & Weissman, S.M. *Cell* **26**, 141 (1981).

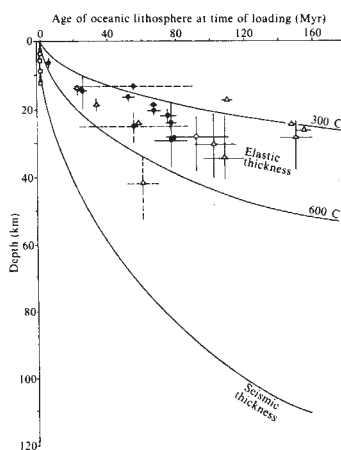
A.B. Watts is Professor of Geology in the Department of Geological Sciences, Columbia University, Palisades, New York 10964.

oceanic islands have shown that the elastic thickness of oceanic lithosphere, which is significantly smaller than the seismic thickness, increases with the age of the lithosphere at the time of loading. The large range of ages represented by these features (1–56 Myr) suggest that on loading, the lithosphere rapidly relaxes from its initial short-term (seismic?) thickness to a long-term mechanical thickness. Subsequent relaxation has not been observed, but may occur on very long (> 56 Myr) time scales. The smallest values of elastic thickness (5–8 km) have been determined at features which form on young oceanic lithosphere, such as the western Walvis ridge guyots, while the largest values (17–37 km) have been determined at features which form on old oceanic lithosphere, such as the Hawaiian islands. These variations suggest⁸ that as the oceanic lithosphere cools, it becomes progressively more rigid in its response to volcanic loads.

The flexure model, deduced from these studies, is referred to as 'elastic', even though the model parameters actually vary with time. The elastic thickness apparently changes rapidly following loading, approaching an asymptotic value that depends on the plate age. The lithosphere only appears elastic, therefore, after some time has elapsed (~1 Myr) and, even then, the thickness of the elastic layer is determined by the plate age.

The actual long-term mechanical properties of the lithosphere are, of course, likely to be more complex than would be predicted by an elastic model. A difficulty is that flexure studies only reveal information on the average mechanical properties of the lithosphere. The estimation of elastic thickness from gravity and geoid data, for example, is complicated by uncertainties in the volcanic load and infill densities, the crustal structure before flexure, and the lateral and vertical extents of lithospheric heating at the time of loading. But as Lambeck⁹ and others have pointed out, gravity and geoid data are consistent with an elastic model in which the elastic thickness increases with age of the lithosphere at the time of loading. A more difficult problem, though, is that there is no maximum stress a purely elastic plate can store; yet the actual materials of the lithosphere are likely to have some ultimate strength. Goetze and Evans¹⁰ have pointed out, however, that if a more realistic 'yield stress envelope' is used for the lithosphere (based on data from experimental rock mechanics), an elastic core would be expected to develop during flexure that corresponds closely to the elastic thickness determined using the elastic model.

Although an elastic model has been widely used in flexure studies^{4–9}, it has been criticized by some workers who prefer a viscoelastic (Maxwell) model. A viscoelastic model is characterized by an initial elastic response to a load, followed



Estimates of the elastic thickness (or mechanical thickness) of oceanic lithosphere plotted against age of the lithosphere at the time of loading. The solid symbols are estimates from studies by different workers at seamounts and oceanic islands, open symbols are estimates from studies at mid-ocean ridges (squares), river deltas (diamond) and deep sea trench-outer rises (triangles). As the age of most of the features range from a few Myr to several tens of Myr, flexure studies indicate that on loading the mechanically supportive part of the lithosphere relaxes from its short-term (seismic) thickness to a long-term mechanical thickness. The long-term thickness does not appear to change appreciably with time. Thus, flexure is likely to be an important controlling mechanism in the tectonic evolution of geological features in the oceans and continents.

by a viscous relaxation that rapidly increases with age of the load. Since the model incorporates a form of stress relaxation, the 'equivalent' elastic thickness would be expected to decrease as the load increases in age, the decrease being most apparent for long times (compared with the Maxwell relaxation time) and large loads. But elastic and viscoelastic models produce flexure curves of similar shape. Thus, the best way to distinguish between them is to determine whether the elastic thickness is controlled by the age of the load (viscoelastic model) or by the age of the plate (elastic model).

Walcott⁴ first suggested, based on the results of flexure studies of continents and oceans, that the elastic thickness of the lithosphere depends on the age of the load, rather than on the plate age. A difficulty with his study, however, was that he did not satisfactorily account for differences in size or tectonic setting of the various geological features used in these studies. Recently, Lambeck¹¹ accounted for these differences in a study of the Society and Cook Islands in the South Pacific Ocean and argued for a dependence of the elastic thickness on age of these loads, that indicated a Maxwell time of about 10^7 yr for the oceanic lithosphere. The differences in age between the Society and Cook Islands are too small (1.4–3.1 Myr), however, to resolve

satisfactorily whether there has been more relaxation beneath the older load (Cook) than the younger load (Society). The best example of a dependence of the elastic thickness on age of load occurs, in fact, along the Hawaiian–Emperor seamount chain; the younger Hawaiian ridge (4–18 Myr) shows the larger elastic thickness (37 km) and the older Emperor seamounts (44–58.5 Myr) the smaller thickness (22 km). But as Watts⁸ pointed out, the small values for the Emperor seamounts can be more simply explained by an elastic, rather than a viscoelastic, model since these seamounts were originally emplaced on young, weak, oceanic lithosphere.

This discussion suggests, therefore, that a consensus is emerging from flexure studies at seamounts and oceanic islands favouring the elastic model as the most useful working hypothesis for the long-term mechanical properties of the lithosphere. The model can also explain (with some modification) the topography of the outer rise seaward of deep-sea trenches, the deep structure of large river deltas, the pattern of late glacial rebound due to the receding Laurentide ice caps and the patterns of gravity anomalies over orogenic belts and Precambrian granite plutons. The model is therefore currently enjoying a wide range of applications in geology that include the determination of the tectonic setting (on-ridge or off-ridge) of intraplate volcanoes, the nature of the control on stratigraphical sequences in sedimentary basins and the gravity and geoid effect of deep processes beneath the lithosphere, such as layered mantle convection.

Clearly, the elastic model remains a working hypothesis that needs to be tested in the future. A critical test would be to determine the actual seismic thickness of the crust and lithosphere in the region of a large volcanic load, such as the Hawaiian islands. Recently, two-ship multichannel seismic experiments, using long arrays and repetitive sound sources, have been carried out in the western and eastern Pacific Ocean¹² that have revealed nearly continuous reflections from the 'Moho' discontinuity at the base of the oceanic crust. Similar experiments in the central Pacific, near the Hawaiian islands, offer the most promise, therefore, to test the flexure model of isostasy during the next decade. □

- Hayford, J.F. & Bowie, W. *Spec. Publ.* 10 (Coast and Geodetic Survey, Washington DC, 1912).
- Vening Meinesz, F.A. *Publ. Neth. Geodetic commun.* (Delftsche Uitgevers Maatschappij, Delft, 1948).
- Hertz, H. *Wiedemann's Annalen* 22, 449 (1884).
- Walcott, R.I. *J. geophys. Res.* 75, 3941 (1970).
- Watts, A.B. & Cochran, J.R. *Geophys. J.* 38, 119 (1974).
- McNutt, M. & Menard, H.W. *J. geophys. Res.* 83, 1206 (1978).
- Roufosse, M.C. *Proc. 9th GEOP. Conf. Dept Geodetic Sci. Rep.* 280, 261 (1978).
- Watts, A.B. *J. geophys. Res.* 83, 5989 (1978).
- Cazenave, A., Lago, B., Dominti, K. & Lambeck, K. *Geophys. J.* 63, 233 (1980).
- Goetze, C. & Evans, B. *Geophys. J.* 59, 463 (1979).
- Lambeck, K. *Earth planet. Sci. Lett.* 55, 482 (1981).
- Stoffa, P.L., Buhl, P., Herron, T.J. & Ludwig, W.J. *Mar. Geol.* 35, 83 (1980).

Gene rearrangement and the generation of diversity

from Miranda Robertson

THE eukaryotic genome has recently begun to seem less stable, both in evolution¹ and in ontogeny², than many biologists had assumed. These trends prompted the Institute of Genetics to invite to its annual Spring meeting speakers on six genetically labile systems: the MHC locus and the immunoglobulin genes of mammals, the genes encoding the surface antigens of trypanosomes, and the genes of plant and animal tumour cells. The MHC locus, which illustrates the evolutionary lability of the vertebrate genome, will be discussed in detail in a later article. What follows here is concerned with the generation of diversity on a much shorter time-scale, in antibodies during the lifetime of vertebrates, and in the surface antigens of parasitic trypanosomes in the course of an infection.*

Both the antibodies of vertebrates and the surface antigens of trypanosomes have evolved the means of extremely rapid diversification through DNA rearrangements leading directly to changes in gene expression. But the selective pressures on antibody-producing vertebrate lymphocytes are different in fundamental ways from those acting on trypanosomes; and those differences may be reflected in the mechanisms by which the genetic changes are achieved. Speakers at Cologne explored both the causes and the consequences of variation in these two systems — and touched on some implications for the normal and abnormal development of multicellular animals.

Deletion and recombination in immunoglobulin

The somatic rearrangement of the DNA of antibody genes takes place in lymphocytes of the lineage whose end-product is the differentiated antibody-secreting B cell, or plasma cell. Its effect is to join the two normally separate gene segments (the V and J segments) to encode the variable region of the light chains of the antibody molecule, and of three separate gene segments (the V, D and J segments) to encode the variable region of the heavy chain (see summary diagram in ref. 2). It is no longer disputed that the variability of the variable regions of antibodies arises from the largely random assembly of each of the coding sequences from a large pool of germ-line V segments, a smaller pool of J segments, and (for heavy-chain V regions) a pool of D segments whose extent is still unknown. There is however vigorous debate about the mechanisms by which the recombination takes place.

The only obvious clue yielded by the recombining DNA itself is the presence of conserved 'signal' sequences flanking each of the separate gene segments (see legend to Figure 1). Because these sequences contain inverted DNA repeats, Leder³ and Tonegawa⁴, both in 1979 working on the light-chain genes of mice, proposed that the V-J join results from the formation of a

stem structure through intrachromatid base-pairing in which the DNA intervening between the V and the J loops out and is deleted (Fig. 1a). In support of this proposal, they reported the absence of the intervening DNA in immunoglobulin-secreting myeloma cells.

The deletion theory however very soon encountered a challenge in the work of Steinmetz *et al.*⁵, who found in a mouse myeloma a fragment of the deleted DNA containing a recognizable pair of back-to-back signal heptanucleotides. It is clear from Fig. 1a that such a configuration could be generated by the excision of the Leder/Tonegawa loop-and-stem if it were

ligated at its base to form a circle. But the persistence of such circles in antibody-secreting cells is not consistent with the disappearance of the intervening DNA from the cells analyzed by Leder and Tonegawa.

Furthermore, the persistence of back-to-back signal sequences is quite general: they have now been found in many more myelomas and hybridomas, both by Walfield *et al.*⁶ and by Van Ness and colleagues (M. Weigert, Institute of Cancer Research, Philadelphia). It has thus become necessary to think again about the mechanism of V-J joining; and one alternative, advocated at Cologne by Weigert and by J. Höchtel (from the laboratory of H. Zachau at the University of Munich where the first back-to-back signal fragment was found) is mitotic recombination resulting in the unequal exchange of DNA between sister chromatids (Fig. 1b).

The essential difference between their theory and the Leder/Tonegawa theory is that while the loop-and-stem mechanism

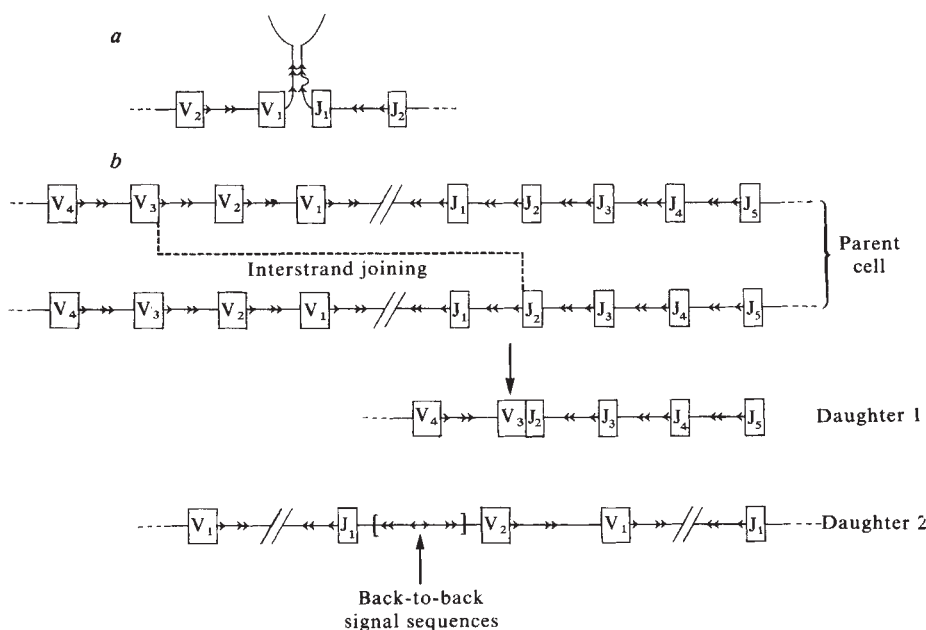


Fig 1. The joining signals for the V and J segments of light-chain genes are a conserved heptanucleotide (>) followed by a conserved nonanucleotide (>), separated by about twelve base-pairs of non-conserved DNA in the case of the V-segment signal, and about 24 base-pairs in the J-segment signal.

For the heavy chain, which is not illustrated here, the conserved sequences are the same but both V and J signals have a 24-base-pair non-conserved spacer. The D segments, which in the assembled coding region come between them, are flanked on both sides by signals with a 12-base-pair spacer. This arrangement has led to the enunciation of the so-called 12-24 rule for segment joining: joining is held to occur only between signals with different spacer lengths.

(a) Hypothetical loop-and-stem structure giving rise to a V-J join. If the loop were ligated at the base of the stem, the result would be a back-to-back pair of signal heptanucleotides.

(b) The consequences of mitotic recombination with an unequal exchange of material between sister chromatids. One daughter inherits a chromosome with a functional V-J join, the other a back-to-back heptanucleotide pair and the potential for a future V-J join. Only one pair of sister chromatids is shown for the dividing parent, and only one chromosome of each daughter: diploid cells will, of course, have two.

Miranda Robertson is associate editor of Nature.

*The Spring Meeting on Molecular Biology, whose theme was the Genetic Basis of Diversity, was held in Cologne on February 24-27, 1982.

entails the redistribution of DNA within a chromatid, interstrand joining between sister chromatids entails redistribution between chromatids, with subsequent segregation of the two chromatids to different daughter cells.

Höchtel and Weigert presented circumstantial evidence for such segregation at Cologne. Höchtel *et al.*⁷ have now sequenced back-to-back signal fragments from three myelomas, and they find that the flanking regions of the back-to-back signals are not those of the V and J segments actually rearranged in the myeloma cells. This can easily be explained if the myelomas had inherited a chromosome such as that illustrated for daughter 2 in Fig. 1b, productive rearrangement of the expressed segments occurring either by a second recombination involving the same chromosome, or by a rearrangement on the other chromosome.

Mitotic recombination can also effortlessly explain the 'deletion' of DNA seen by Leder and Tonegawa, and by Weigert in some cell lines, since after two or more rounds of mitotic recombination a proportion of cells will inherit two shortened chromosomes. Equally, however, it follows that the total DNA content of a population of normal B lymphocytes should be conserved. This prediction has been tested by Van Ness *et al.*⁸ who have established that most of the DNA upstream of the J region is retained in a population of normal B lymphocytes from mouse spleen.

Multiple rearrangements

But there is one feature of the data on V-J joining whose explanation is not so obvious. That is the preponderance of rearrangements involving J₁. This is reflected both in the proteins produced by myelomas and in the sequenced back-to-back fragments, all of which so far prove to contain flanking DNA from the J₁ segment. Weigert suggests that this may be a consequence of rearrangements of upstream J segments remaining on the chromosome containing the back-to-back signal fragment: the chromosome of daughter 2 in Fig. 1b illustrates such an upstream J. The downstream J segments remaining on daughter 1 would never be rearranged because rearrangement ceases once a V segment has been productively joined to a J. This would lead to an overall bias in favour of J₁.

On the other hand, if multiple rearrangements are frequent then the back-to-back signal fragments would no longer be expected to correspond in a one-to-one fashion with the productively rearranged segments in a given cell, and one of the stronger arguments for exchange between sister chromatids would be weakened. Indeed Baltimore, who has evidence that immunoglobulin genes are capable of multiple rearrangements, places quite a different interpretation on the back-to-back signal sequences that he, too, finds associated with rearranged light-chain

genes. His evidence for multiple rearrangements comes, not from light-chain genes but from a recent analysis of heavy-chain V-D-J joining by Alt and Baltimore⁹.

By following the fate of the D segments in a pre-B cell line transformed by murine leukaemia virus (MuLV) and still undergoing DNA rearrangement, he and his colleagues have been able to show that multiple rearrangements of D segments are possible; in one case three separate joining events between D and J segments alone must have occurred on a single chromosome. The final result of these rearrangements is a D segment sandwiched between two J segments and therefore incapable of participating in heavy-chain production. But two features of the

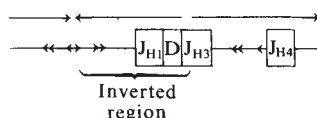


Fig. 2. Abortively rearranged heavy-chain DNA of a virus-transformed cell line described by Alt and Baltimore⁹, the arrows indicate the relative orientation of the DNA.

rearranged chromosome have important general implications for the mechanism of immunoglobulin-gene rearrangement. First, a pair of back-to-back heptamers has been generated upstream from the joined segments; and second, the DNA between the heptamers and the downstream margin of the D segment is inverted (Fig. 2). While the heptamers could in principle be explained by recombination between sister chromatids, the inversion cannot, since it would entail breaking the chromosome. Baltimore believes that the back-to-back heptamers in rearranged light-chain DNA, too, may be generated by the inversion of V segments during V-J joining. (The orientation of the V segments relative to the J segments in germ-line DNA is not in fact known.) At this stage, however, no mechanism for V-J joining can be definitively ruled out.

More diversity by mutation and conversion

In any case there is more diversity in the variable-region coding sequences than either mitotic recombination or intrachromosomal deletion can explain on its own. It turns out that somatic mutation occurs perhaps during and certainly after the DNA rearrangements that join the different gene segments^{2,10}; this can result in a very large number of substitutions — as many as 44 in a single V region. L. Hood (California Institute of Technology), on the basis of a detailed analysis of expressed and germ-line sequences encoding the V regions of antibodies that bind the hapten phosphorylcholine (PC), has concluded that there must be a special mechanism for generating localized mutations, since

almost all of them are confined to the DNA encoding the V region and its immediate vicinity¹¹. He calls the phenomenon hypermutation.

It can, however, be difficult to be certain that the substitutions are due to mutation and not recombination between members of the extensive V-gene family. Hood's evidence rests on the identification of a family of germ-line V segments that hybridizes strongly with the expressed anti-PC V sequences and not with other germ-line genes; and furthermore he finds substitutions in the J region, for which all the germ-line sequences are known.

The danger of postulating mutation in the absence of complete information on germ-line sequences was illustrated by collaborators of K. Rajewsky and K. Bayreuther (University of Cologne), who in similar studies have discovered a V-region variant arising *in vitro* in which what was apparently a cluster of mutations in the heavy-chain V region proved to be identical to a part of the germ-line sequence of a different V segment. They have interpreted this striking discovery as a possible instance of gene conversion — a mode of recombination increasingly invoked to account for otherwise inexplicable blocks of homology.

The question of whether conversion makes a significant contribution to antibody diversity *in vivo* however remains open.

Structure and function of antibodies

The random recombination and subsequent mutation, or conversion, of immunoglobulin variable-region gene segments guarantee diversity of structure, but they by no means guarantee diversity of function. The extremely ill understood relationship between the two has been the chief preoccupation of S. Rudikoff (National Institutes of Health) who reported on the basis of an analysis of variants of anti-PC antibodies, that amino-acid substitutions may make remarkably little radical difference to antigen-binding specificity. Hood hazarded that the choice of germ-line segments probably broadly determines the character of the antigen bound, while mutations (or conversions) 'fine-tune' the binding properties of the antibody.

An extraordinarily elegant illustration of this principle was presented by W. Gerhard (Wistar Institute, Philadelphia), in whose laboratory a large panel of hybridomas has been constructed from individual inbred mice immunized with influenza A virus, and the monoclonal antibodies tested for binding to a range of mutant viral haemagglutinins. From among these antibodies, one group could be identified, on the basis of restriction mapping, as a family probably derived from the same rearrangement of germ-line segments, and thus presumably from the same precursor B cell. The antigen-binding properties of most of them could, however, be clearly distinguished by single-point mutant

haemagglutinins, each of which was recognized by only two or three of the antibodies.

What this means, in practical terms, is that somatic mutation in memory B cells should enable an immunized mammal to keep pace with the continuous minor antigenic drift of the influenza haemagglutinin molecule. It may also explain why, by contrast, the more dramatic antigenic shift that occurs every few years can give rise to pandemics¹²: the more profound structural effects of the shift may demand a fresh combination of germ-line segments and leave the animal temporarily unprotected.

Recombination and expression of genes for trypanosomes surface antigens

Like the influenza virus, the African trypanosome has evolved the ability to change its antigenic coat proteins and thus evade the immune armoury of its vertebrate host. Indeed, the surface antigens of trypanosomes change sequentially in the course of a single infection and generally outflank the immune system altogether. But the parallel between trypanosomes and influenza viruses is in many ways less striking than the parallel between trypanosomes and the B lymphocytes of their victims. Trypanosomes, like B lymphocytes, have a family of related genes encoding surface proteins; and, like B lymphocytes, they can vary the expression of the genes by rearrangements of the chromosome containing them¹³.

The manner of the rearrangement may however be less like that of antibody genes than that of the mating-type genes in yeast, since the expression of a variable surface glycoprotein (VSG) gene in trypanosomes seems to be preceded, at least in some cases, by duplication and transposition¹³, whereas no duplication (so far as is known) accompanies the rearrangement of antibody V segments. P. Borst (University of Amsterdam) has suggested that the duplication-transposition of trypanosome VSG genes could be another instance of gene conversion, in which the gene in the expressed locus — the so-called expression-linked copy — is converted sequentially by the other, co-called basic copy, genes. (The conversion event postulated by the Cologne group in immunoglobulin V segments occurs after gene rearrangement and not as part of its mechanism.) There is moreover no particular reason to expect that the same recombination mechanism will operate in B lymphocytes, which do not need to maintain their genomes intact for transmission to the next generation, and trypanosomes, which do.

Transposition and activation of genes

A significant point of similarity however is that in both cases the rearrangement of DNA results in the activation of the rearranged gene. The question is, how? In the case of antibody genes, it is known that there are promoters upstream of not only the unrearranged constant-region

segments at the transcribed locus but each of the germ-line V segments as well. The constant-region sequences are transcribed at low levels from unrearranged chromosomes in both B and T lymphocytes, suggesting that the locus is constitutively active in the lineage. The unrearranged V segments by contrast are silent. Van Ness *et al.* have established that recombination of the V segment (and its promoter) with the J and C segments increases the transcriptional activity of the C region severalfold¹³. They conclude that the total transcriptional activity may depend on the combined effects of both V segment promoters, and the structure of the chromatin at the site of the C segment — the appropriate chromatin structure being the necessary condition for expression. (There is in fact evidence for a region of DNase 1 sensitivity upstream of the C-region genes¹⁵.)

Similar factors may determine VSG gene expression in trypanosomes, though Borst believes the powerful promoter may be at the expression site rather than at the basic copy site. He reported at Cologne that the mRNA transcribed from the expressed gene contains a 5' sequence that does not correspond to the 5' flank of the basic-copy gene, and suggested that it is derived from the DNA at the expressed locus, which may contain an upstream promoter¹⁶. This hypothesis could clearly be tested by cloning the DNA of the expression-linked copy.

But trypanosomes in have in any case more than one way of changing the expression of their VSG genes. Williams *et al.*¹⁷ have found rearrangement of the expressed gene without evidence of duplication; and S. Longacre (Pasteur Institute, Paris) reported that duplicated genes in *T. equiperdum* may persist unexpressed while a second duplicated gene in a different locus is transcribed.

If the rearrangement of trypanosome VSG genes does not bring them under the control of a promoter, perhaps, like

antibody genes, they bring their own promoters with them and activation is due to the state of the surrounding chromatin.

This is the mechanism suggested by Nasmyth *et al.*¹⁸ to account for the alternating expression of the α and α mating-type genes of yeast. Each yeast cell carries one silent copy of each of the two related genes, a duplicated α or α gene being expressed in alternate generations at the MAT locus. There is some evidence that the chromatin at the MAT locus is more loosely packed than at the silent loci.

But the importance of chromatin structure may not be confined to its effect on transcriptional activity. Mating-type genes, VSG genes, and immunoglobulin V segments are not transposed to random sites in the genome, and the state of the chromatin may help to determine where they recombine. In every case, new sequences are transposed to an already active site, and Van Ness *et al.* have suggested that, in the case of antibody genes, transcriptional activity may be a prerequisite for recombination. In this context, it may be extremely significant that in mutant yeast strains in which the normally silent mating-type loci are transcribed, these loci, like the normally active site, switch between α and α sequences; in wild-type strains, these silent loci are never expressed and do not switch¹⁹.

Transformation and differentiation

Are these variations in position and expression of related genes a peculiarity of a handful of highly specialized genetic systems, or have they (as Strathern and Herskowitz have suggested for yeast²⁰) important implications for the study of differentiation in general? It has already been pointed out²¹ that the existence of three families of mobile promoters on the chromosomes containing the three pools of V segments poses a serious oncogenic threat. Since it is now generally held that cancer may be caused by normal cellular genes promoted to abnormally high levels of activity, the arrival of a V segment complete with promoter in the vicinity of such a gene might lead to cancer; and indeed there is evidence that some leukaemias may arise in just this way²¹.

Speculations on the part played by labile gene families in normal development however are so far largely unconstrained by evidence. Thus unfettered, Hood launched a pan-explanatory schema in which embryogenesis is guided by varying surface markers differentially and sequentially expressed on cells with different developmental destinies, each additional surface marker further narrowing the differentiation pathway open to its bearer — just as differential expression of antibody V-regions narrows the interactions open to a lymphocyte. Such is the present pace of progress that today's imaginative leap may land in tomorrow's text-book. □

1. see Jeffreys, A.J. & Harris, S. *Nature* **296**, 9 (1982).
2. Robertson, M. *Nature* **290**, 625 (1981).
3. Max, E.E., Seidman, J.C. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3540 (1979).
4. Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. *Nature* **290**, 288 (1979).
5. Steinmetz, M., Altenburger, W. & Zachau, H.G. *Nucleic Acids Res.* **8**, 1709 (1980).
6. Walfield, A., Selsing, E., Arp, B. & Storb, U. *Nucleic Acids Res.* **9**, 1101 (1981).
7. Hocht, H.J., Müller, C.R. & Zachau, H.G. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1383 (1982).
8. Van Ness, B.G., Coleclough, C., Perry, R.P. & Weigert, M. *Proc. natn. Acad. Sci. U.S.A.* (1982).
9. Alt, F. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.*, in the press.
10. Alt, F. *et al. Cell* **27**, 381 (1981).
11. Kim, S., Davis, M., Potter, P. & Hood, L. *Cell* **27**, 573 (1981).
12. Webster, R.G., Laver, W.G., Air, G.M. & Schild, G.C. *Nature* **296**, 115 (1982).
13. Hoefjmakers, J.H.J. *et al. Nature* **284**, 78 (1980).
14. Van Ness, B.G. *et al. Cell* **27**, 593 (1981).
15. Storb, U., Arp, B. & Wilson, R. *Nature* **294**, 90 (1981).
16. Bernard, A. *et al. Cell* **27**, 497 (1981).
17. Williams, R.O., Young, J.R. & Majwa, P.A.O. *Nature* **282**, 847 (1979).
18. Nasmyth, K. *et al. Nature* **289**, 244 (1981).
19. Klar, A.J.S., Strathern, J.N. & Hicks, J.B. *Cell* **25**, 517 (1981).
20. Strathern, J.N. & Herskowitz, I. *Cell* **17**, 371 (1979).
21. Klein, G. *Nature* **294**, 313 (1981).

REVIEW ARTICLE

Chemistry of hot springs on the East Pacific Rise and their effluent dispersal

J. M. Edmond*, K. L. Von Damm, R. E. McDuff & C. I. Measures

Department of Earth and Planetary Sciences, Massachusetts Institute of Technology, E34-201, Cambridge, Massachusetts 02139, USA

Cold groundwater circulating through submarine ridge crests reacts with hot basalts at temperatures in excess of 300 °C. The resulting solutions ascend and emerge as hot springs on the sea floor. This ridge crest hydrothermal activity is pervasive and chemically uniform. The fluxes of the various elements injected into, or removed from, the ocean are substantial compared with fluvial transport. The dispersal of the effluent is shown to be controlled by global circulation at mid-depth.

LARGE-SCALE, high temperature, hydrothermal activity is a ubiquitous concomitant to the production of oceanic crust at submarine ridge crests. The establishment of this fact has been one of the most important recent developments in oceanography and, indeed, in the earth sciences as a whole. Molten rock is injected at the centres of sea floor spreading at temperatures of around 1,200 °C. The pre-existing crust is highly permeable because of faulting and of fissuring induced by thermal contraction. Cold 'groundwater' derived from the overlying ocean can thus circulate through the crust and penetrates the zone of active intrusion along cracking fronts¹. Here reactions take place between the seawater and the hot basalt at temperatures in excess of 300 °C. The resulting solutions ascend to the sea floor where they emerge as hot springs.

Geological evidence

The phenomenon was first proposed by Elder² from analogy with geothermal systems associated with continental igneous activity. At about the same time Bostrom and Peterson³ invoked ridge crest hot springs of undefined scale as the source of the elevated concentrations of iron and manganese which they had shown to be a characteristic feature of the sediments on the crestal zone of the East Pacific Rise. Such a local source was required by their observations since abyssal sediments remote from the axes showed no such enrichments. As data were accumulated on the chemistry and mineralogy of basalts dredged from the ridge crests it became clear that, in many cases, they had been pervasively altered by reaction with seawater^{4,5}. The partitioning of oxygen isotopes between coexisting secondary minerals, the reaction products of this alteration, demonstrated that the temperatures of the interactions ranged up to several hundred degrees centigrade⁶.

Parallel investigations in ophiolites, slabs of oceanic crust tectonically emplaced onto the continents, gave similar results and established that seawater must penetrate over 5 km into the intrusion zone at the ridge crests⁷. Metalliferous sediments were found to be generally associated with the uppermost pillow lavas in the ophiolite sections and to be of hydrothermal origin^{8,9}. Ore deposits composed of massive accumulations of pyrite (FeS₂) in the same terrains were shown to require very large volumes of high temperature solutions for their transport and emplacement¹⁰.

Results from the Deep Sea Drilling Project demonstrated that metalliferous sediments are the usual basal deposit in the contemporary oceanic sediment column¹¹. The underlying basalts, where sampled, showed evidence of reaction with seawater over a wide range of temperatures¹². Detection of the primordial, non-radiogenic, isotope of helium, ³He, at mid-depth in the Pacific water column¹³ provided a strong additional indication that hydrothermal activity is occurring along the crest of the East Pacific Rise.

Quantitative estimates of the global scale of this seawater convection began to emerge in the mid-1970s from comparison of the measured conductive heat loss from young oceanic crust with that predicted from the thermal models of the plates¹⁴. A pronounced negative anomaly was observed everywhere¹⁵. The total deficit for all the active ridges is estimated at between 2 and 8×10^{19} cal yr⁻¹ (ref. 16). Assuming that this heat was removed by convective flow at typical temperatures of 300 °C, as suggested by the ¹⁸O data⁶, then a volume of water equivalent to the whole ocean must circulate through the high temperature intrusion zone in the ridge axes (that is through the 300 °C isotherm) every 8–10 Myr.

The chemical importance of this phenomenon was illustrated in the earliest experimental work on seawater–basalt interactions at elevated temperatures and sea floor pressures¹⁷. Drastic changes were observed in the compositions of the solutions during the reactions. Magnesium and sulphate were rapidly depleted and potassium, calcium and silica enriched. Simple calculations indicated that ridge crest hydrothermal activity might be a major, hitherto unconsidered, component of the chemical mass balance of the oceans over geological time¹⁵.

Chemistry of ridge crest hot springs

Finding and sampling actual examples of ridge crest hot springs proved to be complex. Although numerous indications of hydrothermal activity had been detected in the water column over the axial zone^{18,19}, it was not until 1977 that a comprehensive exploration programme at the Galapagos Spreading Centre in 86° W in the eastern equatorial Pacific succeeded in locating and collecting water from discrete vents²⁰. This required use of high resolution bathymetric charts, an accurately navigated, large-field, survey camera and the research submersible *Alvin*.

The hot springs at the Galapagos Spreading Centre were found to occur in extensive fields²¹ associated with diverse and exotic fauna²². The exit temperatures were low, in the range 6–20 °C above that of the ambient water column at the axial depth (2.05 °C, 2,500 m). However, systematic investigation of the chemistry of the solutions demonstrated that these low and variable temperatures were actually an artefact of large-scale, sub-surface mixing between a very high temperature component and 'groundwater' indistinguishable in properties from the ambient seawater^{23,24}. Individual chemical elements in solution respond differently to this process. The hydrothermal solutions were cool, contained variable amounts of H₂S (10–160 μmol kg⁻¹) and were mildly acidic²⁵. Elements which do not form insoluble phases in such conditions are unaffected by the mixing process.

The experimental data¹⁷ indicated the complete removal of dissolved magnesium during high temperature seawater–basalt reactions. Assuming that the fluids sampled at Galapagos represented a simple mixture between cold 'groundwater' and a single hydrothermal endmember then the temperature of the latter could be estimated by extrapolating the observed negative

* Present address: Institut de Physique du Globe, Université de Paris VI, France.

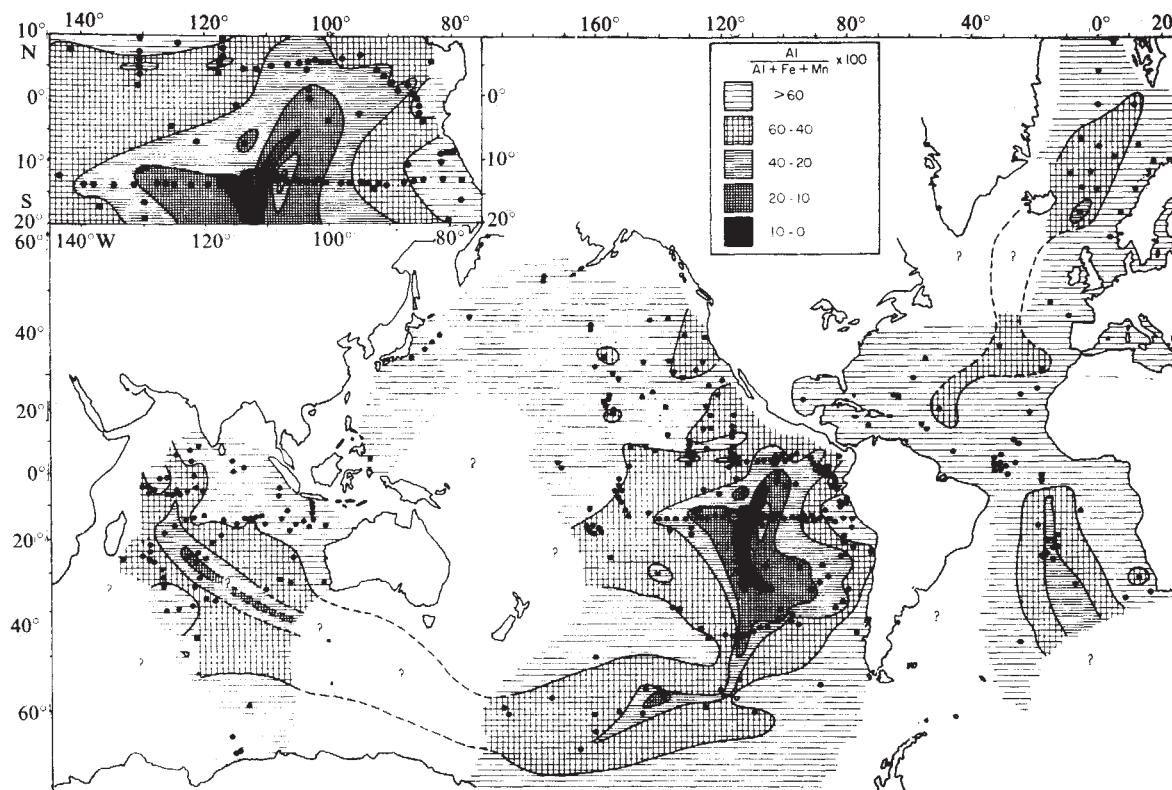


Fig. 1 Distribution of metalliferous surface sediments in the ocean as indicated by the ratio $\text{Al}/(\text{Al}+\text{Fe}+\text{Mn})$. As these deposits are poor in aluminium relative to continental detritus their abundance is inversely proportional to the value of the ratio. Figure reproduced from ref. 38.

magnesium-temperature relationship in the samples to zero concentration. An intercept was obtained at $\sim 350^\circ\text{C}$ (ref. 23). This was confirmed by application of the quartz solubility criterion—quartz geothermometry²⁶—to the parallel data for dissolved silica. Extrapolation of the concentration-temperature trends for the other conservative species then allowed estimates to be made of the composition of this putative 350°C endmember (Table 1). Elements which form insoluble precipitates—sulphides and oxides—in the conditions observed in the Galapagos solutions showed sharp depletions with increasing temperature, their concentration levels extrapolating to zero between 30 and 40°C (ref. 24). Not only were these elements—Cu, Ni, Cd, Se, U, Cr—being removed from the high temperature component of the fluids but also from the admixed groundwaters. It was not possible, therefore, to obtain any information on the concentrations of these potentially ore-forming elements in the high temperature solution.

Despite such complexities some remarkable results were obtained from the Galapagos data: these supported far-reaching conclusions as to the global significance of ridge crest hydrothermal activity. Most important are the fluxes involved. For ^3He the ratio to transported heat is $2.2 \times 10^{-17} \text{ mol cal}^{-1}$ (ref. 27). Integration of the water column ^3He anomaly, coupled with ^{14}C -calibrated models for the oceanic overturn rate, results in a flux of ^3He into the ocean, necessary to maintain the anomaly, of $1,100 \text{ mol yr}^{-1}$ (ref. 28). This is in good agreement with the terrestrial budget which is dominated by a loss term due to gravitational escape to space. If it is assumed that all the ^3He delivered to the oceans is from hot springs in the ratio to heat observed at Galapagos then the associated heat flux can be calculated. This independent estimate agrees exactly with the geophysical results²⁷.

The apparent success of this approach coupled with appreciation of the close constancy in composition of the hydrothermal reactants, tholeiite basalts and seawater, justified similar calculations for the other dissolved species. The resulting values for elemental fluxes into and out of the axial zone are in many cases comparable with or greater than those estimated for fluvial

transport of the products of continental weathering²³. Thus the ridge is the major sink for magnesium and sulphate, the major source for lithium, rubidium and manganese and a substantial contributor of calcium, barium and silica. Several long-standing problems in the construction of oceanic chemical balances are resolved by inclusion of the hydrothermal term. The situation for the non-conservative elements cannot be evaluated²² as the global distribution of exit temperatures—the average extent of sub-surface mixing—is, of course, unknown.

The validity of the deductions based on the Galapagos data was confirmed by the discovery, 2 years later, of hydrothermal fields on the crest of the East Pacific Rise in 21°N , south of the mouth of the Gulf of California^{29,30}. Here, in depths of 2,500 m, solutions at temperatures of 350°C were observed issuing from the sea floor. They had associated with them constructional features, chimneys up to 10 m high, composed

Table 1 Comparison of the estimated composition of a 350°C hydrothermal endmember based on extrapolation of the Galapagos data with that observed at 21°N

	Galapagos	21°N	Seawater
Li ($\mu\text{mol kg}^{-1}$)	1,142–689	820	28
K (mmol kg^{-1})	18.8	25.0	10.1
Rb ($\mu\text{mol kg}^{-1}$)	20.3–13.4	26.0	1.32
Mg (mmol kg^{-1})	0	0	52.7
Ca (mmol kg^{-1})	40.2–24.6	21.5	10.3
Sr ($\mu\text{mol kg}^{-1}$)	87	90	87
Ba ($\mu\text{mol kg}^{-1}$)	42.6–17.2	95–35	0.145
Mn ($\mu\text{mol kg}^{-1}$)	1,140–360	610	0.002
Fe ($\mu\text{mol kg}^{-1}$)	+	1,800	—
Si (mmol kg^{-1})	21.9	21.5	0.160
SO_4 (mmol kg^{-1})	0	0	28.6
H_2S (mmol kg^{-1})	+	6.5	0

The ranges in the Galapagos results derive from the different trends for composition versus heat observed between individual vent fields²³. +, Non-conservative to sub-surface mixing; —, seawater concentration not accurately known.

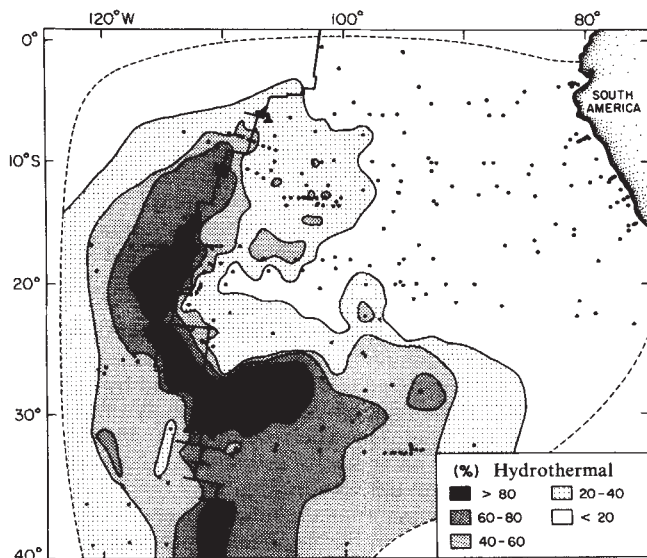


Fig. 2 The weight percentage of the hydrothermal component on a carbonate and salt-free basis in surface sediments of the Nazca Plate³³. The abundance of this component is calculated from chemical data for the bulk sediment using a normative composition model.

predominantly of iron, zinc and copper sulphides. These grow from the hydrothermal solutions as they mix with the ambient seawater. However, most of the precipitation occurs as very fine-grained particles entrained in a turbulent plume rising tens of metres above the orifices of the springs³⁰.

Sampling of these hot springs established that they are indeed the compositional endmember postulated on the basis of the Galapagos data³¹. The waters exit as clear, homogeneous solutions within a few degrees of 350 °C. They are completely depleted in magnesium and sulphate and appear to be saturated with quartz. The concentrations of the conservative species are close to the Galapagos extrapolations (Table 1). The ³He/heat ratio is identical to that from Galapagos³². The validity of the flux calculations is substantially confirmed.

These undiluted solutions also preserve the signatures of elements precipitated during Galapagos-type sub-surface mixing. The iron concentration is 1.8 mmol kg⁻¹, the sulphide 6.5 mmol kg⁻¹. The Fe/Mn ratio is close to 3, coincidentally identical to the average in ridge crest metalliferous sediments³³. This low value, as compared with that in basalt (50–60), indicates the importance of iron oxidation in the reduction of seawater sulphate in the reaction zone. There is also a possible solubility control by pyrrhotite (FeS) although existing thermodynamic data are inadequate to be certain of this³⁴.

The isotopic data as well as confirming the ³He/heat ratio³⁰, show that the substantial enrichment of CO₂ in the solutions is of mantle origin ($\delta^{13}\text{C} = -6\%$). The ¹⁸O in the existing fluids³², +1.6‰ reflects basalt equilibrium, confirming the previous results for dredged rocks and ophiolites. The isotopic composition of the dissolved strontium (ref. 35 and R.E.McD. and J.M.E., unpublished data) is essentially the same as in the basalts, 0.703. It has been realized for some time^{36,37} that a large flux of non-radiogenic strontium is required to balance the river input (>0.712) and maintain the oceanic value at 0.709. The ridge component is sufficient to close this balance, resolving the problem and supporting the generality of the flux calculations.

Dispersal of hydrothermal effluent

The hot springs at 21° N are ore-transporting solutions such as are responsible for the massive sulphide deposits found in ophiolites and greenstone belts¹⁰. However, the abundance of springs at 21° N is apparently insufficient to produce more than small accumulations^{30,31}. In addition their position in the local

topography and hence in the ambient current field prevents significant local deposition of sedimentary sulphides from the plumes. The particulate sulphides and the large amounts of dissolved, divalent manganese are therefore dissipated by the general mid-depth circulation, and on oxidation in the water column precipitate out regionally contributing to the metaliferous sediments characteristic of the East Pacific Rise. During their chemical evolution in the plumes, the water column and sediments these Fe–Mn oxides must interact extensively with other elements in solution producing the diverse compositional and isotopic signatures observed¹¹.

While the chemistry involved in this evolution is not understood beyond the descriptive level, a significant body of information does exist on the distribution of these hydrothermal sediments in the Pacific. The original maps of Bostrom *et al.*³⁸ showed a pronounced asymmetry in the index Al/(Al + Fe + Mn) about the axis of the East Pacific Rise (Fig. 1). The magnitude of this ratio is inversely proportional to the abundance of the hydrothermal component which is, like the hot spring solutions and the water column, depleted in aluminium. More detailed work by Dymond³³ on the surface sediments of the south-east Pacific (Nazca Plate) confirms the southeastern lobe in Bostrom's distribution, centred on 30° S (Fig. 2). Recent work in the Tiki Basin³⁹, between the East Pacific Rise and the Marquesas, establishes the extent and nature of the western limb along 15° S. In both instances hydrothermal precipitates are being transported for thousands of kilometres before eventual deposition. Moreover the direction of transport is strongly dependent on latitude. The western lobe is exactly coincident with the ³He plume recently identified by Lupton and Craig⁴⁰ (Fig. 3). This remarkable feature appears to originate as an intense ³He maximum over the ridge axis at 15° S and propagates westwards at mid-depth with no discernable 'leakage' to the east.

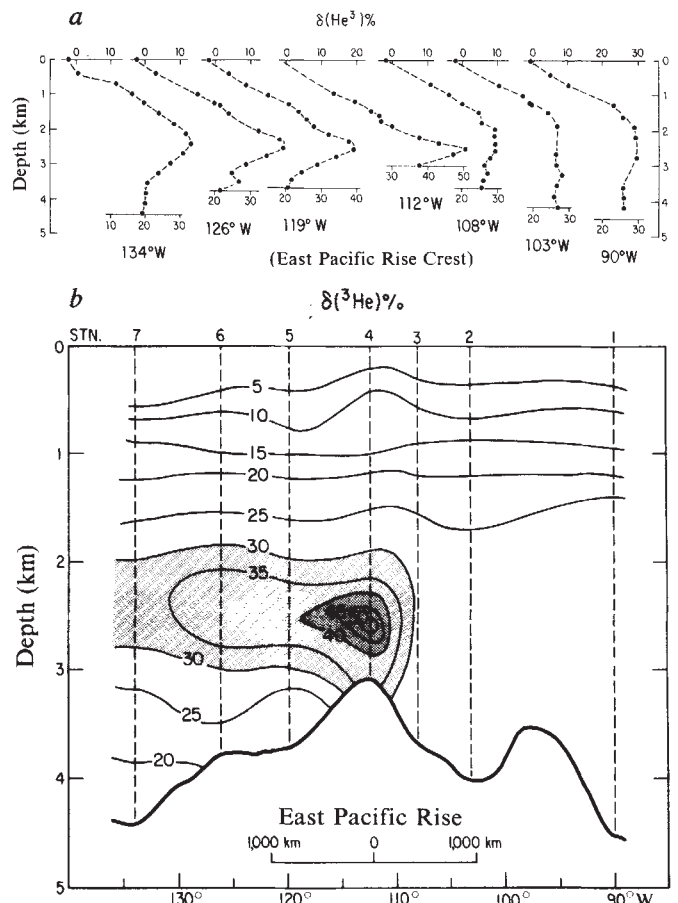


Fig. 3 $\delta^3\text{He}$ section along ~15° S in the south-east Pacific⁴⁰. *a*, The individual profiles; *b*, the contoured values.

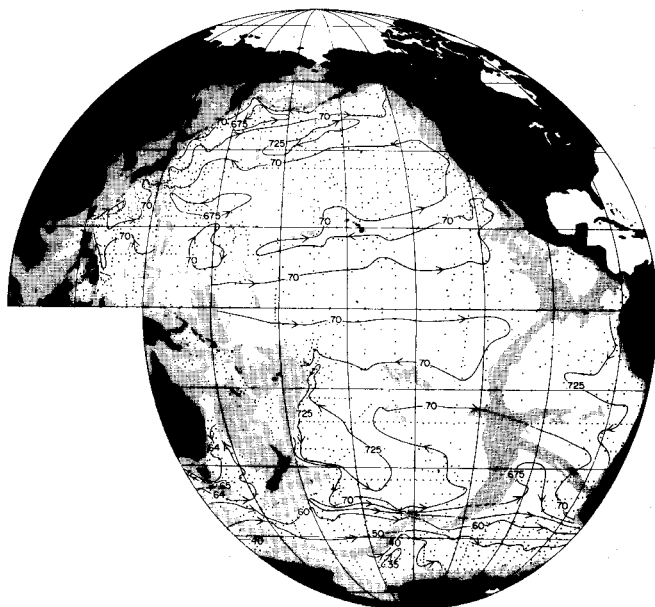


Fig. 4 Steric height at 2,000–3,500 db in dynamic metres ($10 \text{ m}^2 \text{ s}^{-2}$ or 10 J kg^{-1}) for the Pacific⁴². Shaded area is <3,500 m water depth.

These strongly developed asymmetries require a pronounced latitudinal zonality in the mid-depth circulation in the South Pacific, contrary to the predictions of the classical geostrophic models⁴¹. That this indeed exists has been dramatically demonstrated recently by Reid⁴² who mapped the distributions of salinity, oxygen and the major nutrients on an isopycnal surface which is coincident with the ridge crest at mid-latitudes in the Pacific. He has also mapped the steric height at 2,000 db relative to 3,500 db (Fig. 4). Interpreting this measure of the relative geostrophic shear in terms of the mean flow directions (away from the western boundary) results in a circulation scheme consistent with the associated distributions of salinity and the non-conservative species⁴². Moreover comparison of the figures presented here shows the coincidence of the flow field with the distribution of metalliferous sediments and of ^3He to be exact.

The shear field in the South Pacific contains an anti-cyclonic gyre near 35–40°S. An eastwards flow leaves this gyre along 25°S in the east. There is a pronounced westwards flow along 15°S and an eastwards flow in the equatorial region. Flow over the basins east of the East Pacific Rise is generally southwards. This pattern is mirrored in the North Pacific (Fig. 4).

Of particular importance here is the circulation pattern in the region of the axis of the East Pacific Rise. Between about 5° and 15°S the flow is towards the south-southwest parallel to the rise crest. At 15–20°S the flow lines turn abruptly westwards and continue in that direction across the ocean. By 25°S the flow directions are reversed, passing east-southeast directly normal to the axis. In the extreme south-east Pacific, south of about 35°S the flow is to the north-east along the crest.

The exact concordance of the independently derived distributions of hydrothermal sediment and ^3He with the inferred flow is a remarkable demonstration of the importance of the mid-depth circulation in the dispersal of the hot spring effluents injected all along the East Pacific Rise. There are several consequences to this. First, and most obvious, the ^3He 'plume' along 15°S is but one section through a complex three-dimensional feature. The major axial anomaly results from the accumulation of ^3He injected along a flow-line almost 1,000 km in extent—the polewards flowing water above the ridge axis between 5° and 15°S. The plume is the expression of the dissipation of this anomaly along the westwards flow trajectory. The absence of any eastwards 'leakage' of ^3He is as expected. Since the extent of hydrothermal activity under the crestral station on the section (112°W) is not yet known it is not possible to separate the flow-integrating component in the ^3He

maximum from any local injection. The very sharp spike of 50% $\delta^3\text{He}$ (Fig. 3) may reflect the immediate presence of a 21°N-type plume. Recent sampling on the East Pacific Rise has produced isolated ^3He anomalies of even greater magnitudes where the samplers were navigated to be immediately above known hydrothermal vents⁴³. However, the broad maximum in $\delta^3\text{He}$ (35–45%) must be a regional rather than a local feature. Thus the spectacular section along 15°S does not require, of necessity, particularly vigorous hot spring activity on the axial zone beneath it. It does require that there be substantial injection of ^3He along the 5–15°S segment of the East Pacific Rise.

It should be possible to separate local from regional influences on the axial ^3He anomaly using parallel measurements of manganese⁴³. As this element is rapidly removed by oxidation and sedimentation the regional $\text{Mn}/^3\text{He}$ anomaly ratio should be lower than any local one. The resolution of this approach requires much more data for its estimation as the rates of manganese removal are only crudely known¹⁹.

Given the axial alignment of the flow between 5° and 15°S one might expect to see, in this segment, a progressive southwards increase in the abundance of metalliferous sediments accumulating as drift deposits. The data of Dymond³³ indicate this effect. While the spreading rate, a measure of the primary hydrothermal source function, increases from ~15 to 17.5 cm yr^{-1} (ref. 44) between 5° and 15°S the percentage of hydrothermal sediments goes up by over 30% (Fig. 2) and the measured accumulation rates from 100 to >300 $\text{mg cm}^{-2} \text{ kyr}^{-1}$ (ref. 31). While the latter data are sparse and the former biased by independent, systematic, latitudinal changes in the abundance of the other sedimentary components the trend is sufficiently suggestive and consistent to warrant more detailed study.

The systematics observed in the South Pacific have great potential importance in the validation of the estimates of the fluxes of hydrothermal effluents. The problem has two aspects; the generality of the computations for conservative species based on Galapagos and 21°N; the estimation of the general importance of sub-surface mixing—the average exit temperature—for the non-conservative elements. Components from either category which are unreactive in the water column at mid-depth will display an anomaly ratio to ^3He characteristic of the exit values at the hot springs. Sampling, if suitably located in the circulation field (Fig. 4), would provide values for these ratios integrated over particular segments of the ridge axis.

In the water column the concentration of a conservative species, C_A , is the sum of the background concentration, C_B , and an injected concentration, C_I :

$$C_A = C_B + C_I \quad (1)$$

The injected concentration is proportional to the injected ^3He :

$$C_I = R \times ^3\text{He}_I \quad (2)$$

where $R = dC/d^3\text{He}$ for the discharging fluids. In terms of the 350°C endmember:

$$R = \frac{C_{350} - C_A}{^3\text{He}_{350} - ^3\text{He}_A} \quad (3)$$

Combining equations:

$$\frac{C_A - C_B}{C_A} = \left(\frac{C_{350}}{C_A} - 1 \right) \frac{^3\text{He}_A - ^3\text{He}_B}{^3\text{He}_{350} - ^3\text{He}_A} \quad (4)$$

Recognizing that $^3\text{He}_{350} \gg ^3\text{He}_A$ and that $\Delta^3\text{He} = (^3\text{He}_A/^3\text{He}_B) - 1$:

$$\frac{C_A - C_B}{C_A} = \frac{C_I}{C_A} = \left(\frac{C_{350}}{C_A} - 1 \right) \frac{\Delta^3\text{He}}{^3\text{He}_{350}/^3\text{He}_B} \quad (5)$$

When C_{350} is known then the conservative value for C_I/C_A can be specified as a function of the ^3He anomaly: conversely when C_A and C_B are known at a given $\Delta^3\text{He}$ then C_{350} can be calculated.

The value for C_B is important both in determining the magnitude of the water column signature and in allowing accurate computation of the anomaly ratio. For ^3He , C_B is maintained at a low value because of its low solubility and relatively rapid escape from the atmosphere to space. For manganese efficient removal of the oxidized element from the water column produces the same effect. In the simplest case an average C_R can be established from work in the water column. Elements which are non-conservative in the sub-surface mixing regime will have exit concentrations which may be non-linear functions of temperature, depending on the details of the thermodynamic and kinetic responses of the precipitating phases to changing temperature, pH and so on³⁴. This can be established empirically by sampling hot springs over the range of exit temperatures. It will be possible to do this at 21° N where individual hot spring occurrences have been observed spanning the range 20–350 °C. Non-conservative processes in the water column can be examined by sampling along flow lines, for example along 15° S. In principle therefore, the average fluxes from the ridge crest can be determined from work in the water column. However, a considerable amount of analytical development will be required for realizing this objective.

Conclusions

It seems that ridge crest hydrothermal activity is a pervasive and chemically relatively uniform phenomenon. The fluxes of

the various elements injected into (or removed from) the ocean are substantial when compared with fluvial transport. The dispersal of this effluent is controlled by the global oceanic circulation at mid-depth. This effect allows integrated measures of effluent composition to be established from comparison of the water column anomaly ratios to ^3He at locations remote from the ridge axis. The sedimentary record of accumulation of precipitates of hydrothermal origin will therefore provide a history of variations in the circulation itself. This may be of some importance in palaeoclimatic studies. The internal consistency of the water column and sedimentary distributions of ridge crest effluents and the flow field validates the latter in a striking way. Maps of dynamic height for suitable depth ranges are therefore a valuable predictive tool in the planning of disposal of toxic or radioactive waste in the deep sea and in the exploration for and exploitation of mineral deposits of hydrothermal origin. For hot-spring organisms that have motile, benthic larvae it is likely that their dispersal is also controlled by the circulation field. The intriguing possibility therefore exists of a regional biogeography reflecting the latitudinal variations in the flow directions.

We thank Drs Lupton, Dymond and Reid for providing preprints and for stimulating discussions. J.M.E. thanks Dr C. Allègre for providing a place for work. This research is supported at MIT by grants from the NSF and International Decade of Ocean Exploration and from the US Office of Naval Research.

1. Lister, C. R. B. *Geophys. J. R. astr. Soc.* **26**, 515 (1972).
2. Elder, J. W. *Am. geophys. Un. Monogr.* No. 8, 211 (1965).
3. Bostrom, K. & Peterson, M. N. A. *Econ. Geol.* **61**, 1258 (1965).
4. Corliss, J. B. *J. geophys. Res.* **76**, 8128 (1971).
5. Thompson, G. *EOS* **54**, 1015 (1973).
6. Muehlenbachs, K. & Clayton, R. N. *Can. J. Earth Sci.* **9**, 172 (1972).
7. Gregory, R. T. & Taylor, H. P. *J. geophys. Res.* **86**, 2737 (1981).
8. Constantinou, G. & Govett, G. J. S. *Inst. Min. Metal. Trans.* **81**, B34 (1972).
9. Fleet, A. J. & Robertson, A. H. F. *J. geol. Soc. Lond.* **137**, 403 (1980).
10. Constantinou, G. & Govett, G. J. S. *Econ. Geol.* **68**, 843 (1973).
11. Dymond, J. *et al. Bull. geol. Soc. Am.* **84**, 3355 (1973).
12. Muehlenbachs, K. *Init. Rep. DSDP Leg 37* (1976).
13. Clarke, W. B., Beg, M. A. & Craig, H. *Earth planet. Sci. Lett.* **6**, 213 (1969).
14. McKenzie, D. *J. geophys. Res.* **72**, 6261 (1967).
15. Wolery, T. J. & Sleep, N. H. *J. Geol.* **84**, 249 (1975).
16. Sclater, J. G., Parsons, B. & Jaupart, C. *J. geophys. Res.* **86**, 11, 535 (1981).
17. Bischoff, J. & Dickson, F. W. *Earth planet. Sci. Lett.* **25**, 385 (1975).
18. Lupton, J. E., Weiss, R. F. & Craig, H. *Nature* **267**, 603 (1977).
19. Weiss, R. F. *Earth planet. Sci. Lett.* **37**, 257 (1977).
20. Corliss, J. B. *et al. Science* **203**, 1073 (1979).
21. Crane, K. & Ballard, R. D. *J. geophys. Res.* **84**, 6011 (1979).
22. Jannasch, H. W. & Wirsén, C. O. *Bioscience* **29**, 592 (1979).
23. Edmond, *et al. Earth planet. Sci. Lett.* **46**, 1 (1979).
24. Edmond, J. M. *et al. Earth planet. Sci. Lett.* **46**, 19 (1979).
25. Edmond, J. M., Corliss, J. B. & Gordon, L. I. *Proc. 2nd Ewing Symp.* (ed. Talwani, M.) 383 (American Geophysics Union, 1979).
26. Fournier, R. O. & Rowe, J. J. *Am. J. Sci.* **264**, 685 (1966).
27. Jenkins, W. J., Edmond, J. M. & Corliss, J. B. *Nature* **272**, 156 (1978).
28. Craig, H., Clarke, W. B. & Beg, M. A. *Earth planet. Sci. Lett.* **26**, 125 (1975).
29. Cyamex Scientific Team. *Nature* **277**, 523 (1979).
30. Rise Project Group. *Science* **207**, 1421 (1980).
31. Edmond, J. M. *EOS* **61**, 992 (1980).
32. Craig, H., Welhan, J. A., Kim, K., Poreda, R. & Lupton, J. E. *EOS* **61**, 992 (1980).
33. Dymond, J. *Geol. Soc. Am. Mem.* (in the press).
34. McDuff, R. E. & Edmond, J. M. *Earth planet. Sci. Lett.* **57**, 117 (1982).
35. Albarede, F., Michard, A., Minster, J. F. & Michard, G. *TERRA Cog.* **1**, 50 (1981).
36. Brass, G. W. *Geochim. cosmochim. Acta* **40**, 721 (1976).
37. Spooner, E. T. C. *Earth planet. Sci. Lett.* **31**, 167 (1976).
38. Bostrom, K., Peterson, M. N. A., Joensuu, O. & Fisher, D. E. *J. geophys. Res.* **74**, 3261 (1969).
39. Hoffert, M., Karpoff, A. M., Schaaf, A. & Pautot, G. *La Genèse des Nodules de Manganèse* (ed. Lalou, C.) 101 (CNRS, Paris, 1979).
40. Lupton, J. E. & Craig, H. *Science* **214**, 13 (1981).
41. Stommel, H. & Arons, A. B. *Deep Sea Res.* **6**, 217 (1960).
42. Reid, J. L. *Evolution of Physical Oceanography* (eds Wunsch, C. & Warren, B.) 70 (MIT Press, Cambridge, 1981).
43. Merlivat, L., Dimon, B., Boulegue, J. & Francheteau, J. *TERRA Cog.* **1**, R11 (1981).
44. Anderson, R. N., McKenzie, D. & Sclater, J. G. *Earth planet. Sci. Lett.* **18**, 391 (1973).

ARTICLES

Motion of the Local Group of galaxies and isotropy of the Universe

L. Hart* & R. D. Davies

University of Manchester, Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK

H I observations of 84 nearby Sbc spiral galaxies lying in the redshift range 1,000–5,500 km s⁻¹ have been used to determine the motion of the Sun and the Local Group of galaxies relative to the backdrop of galaxies extending out to 70 Mpc. This study yields a Local Group velocity of 436 ± 55 km s⁻¹ towards $l = 264^\circ \pm 18^\circ$, $b = 45^\circ \pm 12^\circ$, in close agreement with the value inferred from the dipole anisotropy of the 3 K cosmic microwave background. The agreement shows that most of the dipole anisotropy is locally (extrinsically) induced and is not an intrinsic property of the Universe as has been implied. The component of the Local Group motion in the direction of the Virgo cluster is 410 ± 55 km s⁻¹, a value similar to that derived recently by independent methods. This implies that the local value of the density of the Universe is 0.15–0.50 of that required to close the Universe.

* Present address: Instituto Argentino de Radioastronomia 1894 Villa Elisa (Prov. de Bs. As), Argentina.

THE dipole anisotropy of amplitude ~ 3.5 mK in the 3 K cosmic microwave background has conventionally been interpreted as the result of the motion of the Sun at a velocity of ~ 400 km s $^{-1}$ towards RA ~ 11.4 h, dec. $\sim 5^\circ$ (refs 1–3). When correction is made for the motion of the Sun relative to the centre of mass of the Local Group of galaxies, the resulting velocity of the Local Group becomes 550 km s $^{-1}$ towards galactic coordinates $l = 265^\circ$, $b = 35^\circ$ (supergalactic coordinates $L = 128^\circ$, $B = -36^\circ$). However, with the available observational data it is by no means certain that the dipole asymmetry in the cosmic microwave background is locally (that is, extrinsically) induced by the motion of the Sun and the Local Group. Dipole and higher order intrinsic asymmetries on the cosmic background have been proposed to explain the microwave observations. For example, Silk and Wilson⁴ predict asymmetries arising from large-scale fluctuations in matter and radiation in the early Universe while Peebles⁵ estimates that gravitational potential fluctuations induced by large-scale clustering could produce the observed large-scale asymmetries.

The suggestion that the microwave dipole asymmetry is largely intrinsic arises from the lack of agreement between the Local Group velocity inferred from the microwave measurements and that obtained from measurements of the motion of the Local Group relative to the backdrop of nearby galaxies. The most widely quoted and systematic search for this motion has been made by Rubin *et al.*^{6,7} who found a motion for the Local Group of 450 km s $^{-1}$ towards $l = 163^\circ$, $b = -11^\circ$, a direction almost orthogonal to that given by the cosmic microwave asymmetry.

A genuine difference between the apparent motion of the Local Group relative to the microwave background and that relative to the backdrop of nearby galaxies has fundamental implications for cosmology. One possible consequence already alluded to is that the microwave background dipole asymmetry may be due to an intrinsic anisotropy^{4,5,8}. Alternatively, if the microwave background is intrinsically isotropic and the dipole asymmetry is the result of the Local Group motion, then the implication is that the whole body of galaxies out to a radius of 90 Mpc is in systematic motion relative to the microwave background with a velocity of 750 km s $^{-1}$. Such large departures from a uniform Hubble flow are not, however, generally accepted⁹. The situation regarding a possible genuine difference between the Local Group motion determined by these two methods has not yet been resolved^{10,11}, particularly as different optical solutions for the Local Group motion have been published elsewhere^{12–14}.

We present here a new approach to deriving the Local Group motion relative to the backdrop of galaxies based on 21-cm wavelength H I observations alone, which obviates many of the difficulties in the optical determinations. Two independent H I data sets each give a Local Group motion in close agreement with that inferred from the dipole anisotropy in the 3K cosmic microwave background. The implications of these results for cosmology are discussed.

Rationale and observational material

The rationale behind our approach to measuring the Local Group motion was to use for distance measurement an H I standard candle independent of the optical methods which gave results disagreeing among themselves. The H I mass (proportional to the integrated H I flux density F_{H}) provides a standard candle and the half power velocity width $\Delta V_{1/2}$ of the integrated H I spectrum provides a third variable which indicates whether the galaxy is a giant or a dwarf. The H I integral is easily measured and calibrated in an all-sky survey and does not require the substantial corrections for galactic absorption, internal absorption, luminosity class, zenith angle, diameter and so on, which are applied to optical data. A detailed discussion of the observations and the selection criteria will be given elsewhere. However, we mention here that the dispersion in the H I mass of the sample of galaxies used in the present work is

similar to the dispersion in the optical luminosity, when these properties are plotted against radius, luminosity class and so on. Ultimately we hope to use, H I, optical and IR data to obtain a composite distance indicator.

The objective of this programme of observations was to obtain a first-order solution for the local group motion using an optimal selection of galaxies whose motions were not seriously influenced by the gravitational potential of the Virgo cluster. Accordingly galaxies were chosen with velocities in the range 1,000–5,500 km s $^{-1}$; the small fraction of these galaxies likely to be most affected by the Virgo cluster were then eliminated from the list. A good sky coverage was achieved for galaxies in each velocity interval.

The galaxies used in this investigation were chosen from the list of 100 spiral galaxies of type Sbc (de Vaucouleurs type $T = 4$) being studied in the $\lambda 21$ -cm line of neutral hydrogen at Jodrell Bank in a programme including optical, radio continuum and X-ray observations. These are essentially all the Sbc galaxies in the Second Reference Catalogue of galaxies¹⁵ (RC2) for which luminosity classes have been assigned and for which the velocity, V_c , corrected for a solar motion of 300 km s $^{-1}$ towards $l = 90^\circ$, $b = 0^\circ$ is $> 1,000$ km s $^{-1}$. Most were of luminosity classes 1 to 4. H I observations have been made by Davies and Johnson (manuscript in preparation) of the integrated H I flux F_{H} , the full velocity width at half-power $\Delta V_{1/2}$ of the integrated profile and the systemic velocity V_c of each galaxy. To obtain a full coverage of the sky, the list was augmented by 12 galaxies at declinations south of dec. = -33° taken from a survey of bright southern galaxies by H. van Woerden *et al.* made at the Australian National Radio Astronomy Observatory (ANRAO), Parkes, New South Wales. The neutral hydrogen flux density scale of the latter survey was tied into the Jodrell Bank scale by comparing 18 galaxies

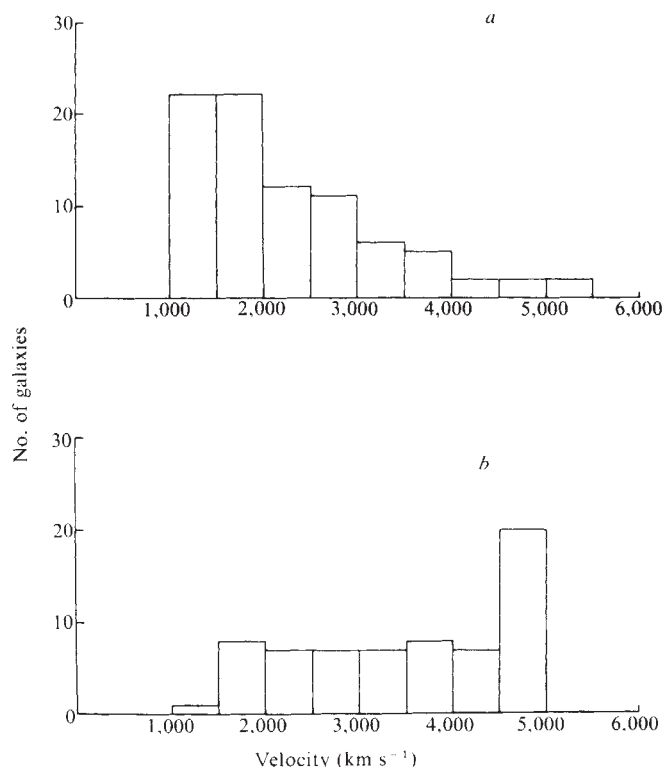


Fig. 1 The velocity distribution of galaxies used in the analysis of the Local Group motion. Velocities have been corrected by the standard solar motion towards the Local Group of 300 km s $^{-1}$ towards $l = 90^\circ$, $b = 0^\circ$. Velocities $< 1,000$ km s $^{-1}$ have not been used. *a*, Sbc ($T = 4$) galaxies from the Jodrell Bank sample supplemented by southern galaxies observed at ANRAO. *b*, Sbc ($T = 5$) galaxies from the Rubin *et al.* sample supplemented by southern galaxies; only galaxies with velocities $< 5,000$ km s $^{-1}$ are used.

Table 1 Observations of the motion of the Sun and the Local Group of galaxies

Method	Velocity of the Sun					Velocity of the Local Group*					No. of galaxies
	$V_{\odot}$ (km s^{-1})	RA (h)	dec. ($^{\circ}$)	l ($^{\circ}$)	b ($^{\circ}$)	V_{LG} (km s^{-1})	l ($^{\circ}$)	b ($^{\circ}$)	L ($^{\circ}$)	B ($^{\circ}$)	
(1) H I fluxes: Sbc galaxies (1,000 < V_c < 5,500 km s^{-1} , $i > 30^{\circ}$)	314 ± 50	12.9 ± 0.7	10 ± 12	309 ± 35	73 ± 12	383 ± 50	256 ± 18	52 ± 16	110 ± 16	-27 ± 14	84
(2) H I fluxes: Sbc galaxies (as for 1, but excluding high Virgo-centric velocities)	310 ± 55	12.3 ± 0.7	30 ± 12	192 ± 60	83 ± 12	436 ± 55	264 ± 18	45 ± 12	119 ± 14	-29 ± 12	78
(3) H I fluxes: Sc galaxies (1,000 < V_c < 5,000 km s^{-1})	419 ± 62	10.2 ± 0.6	19 ± 9	218 ± 15	53 ± 9	580 ± 62	245 ± 18	35 ± 9	111 ± 13	-45 ± 9	53
(4) Optical luminosity: Sbc galaxies	350 ± 40	7.8 ± 1.8	56 ± 16	162 ± 20	31 ± 16	397 ± 90	216 ± 20	27 ± 15	72 ± 28	-55 ± 16	79
(5) Microwave background† (ref. 2)	420 ± 33	11.6 ± 0.2	-12 ± 5	275 ± 7	46 ± 5	653 ± 33	273 ± 6	27 ± 5	140 ± 6	-35 ± 5	
(6) Microwave background† (ref. 1)	401 ± 60	11.2 ± 0.5	19 ± 8	229 ± 19	67 ± 67	567 ± 60	256 ± 9	41 ± 7	117 ± 8	-36 ± 7	
(7) Average of 2 + 3 + 5 + 6‡ (best estimates of motion)	368 ± 50	11.5 ± 0.5	15 ± 12	243 ± 30	68 ± 12	546 ± 70	261 ± 9	39 ± 7	122 ± 9	-35 ± 7	

* The solar motion relative to the Local Group of galaxies is taken to be 300 km s^{-1} towards $l = 90^{\circ}$, $b = 0^{\circ}$.

† The microwave background temperature is assumed to be 2.7 K: velocities are reduced by $\sim 40 \text{ km s}^{-1}$ for 3.0 K.

‡ Methods 2 and 3 are combined with weights 2:1.

common to the two lists. F_H and $\Delta V_{1/2}$ are accurate to better than 10% and V_c was accurate to 20 km s^{-1} or better for all the galaxies. The edge-on velocity width of each velocity profile, $\Delta V_{1/2} \csc i = \Delta V_0$, was estimated using the inclination, i , of each galaxy derived from the axial ratio given in RC2 and an assumed intrinsic thickness-to-diameter ratio of 0.2.

Derivation of motion of the Local Group

This analysis is based on the use of the neutral hydrogen mass of a galaxy as a standard candle and using ΔV_0 as a third variable. The use of ΔV_0 in this analysis is similar to using the Fisher-Tully relation for optical distance determination.

In the analysis we take the observed velocity of a galaxy, V_c , corrected to the centre of the Local Group, to be composed of the following components

$$V_c = V_H + \Delta V_G + V_{\text{pec}}$$

where V_H is the velocity it would have in participating in the universal Hubble flow, ΔV_G is the component in the direction of the galaxy of the velocity of the Local Group relative to the backdrop of galaxies, and V_{pec} is the random velocity of the galaxy. V_H determines the true distance of the galaxy. We are interested to determine, from a series of expressions of this form, the vector V_G giving the velocity of the Local Group relative to the backdrop of Sbc galaxies. An estimate of the true distance of each galaxy, and hence V_H , is provided by its measured (apparent) H I flux density F_H where $F_H = M_H/D^2$, M_H is the H I mass and D is the distance.

Our first analysis follows the approach of Rubin *et al.*, where we express the difference in redshift distance and luminosity distance in terms of an H I Hubble modulus, HM , analogous to their optical HM in which

$$HM = 2 \log V_c - a \log \Delta V_0 + \log F_H \quad (1)$$

where a is the index of the calibration relation $M_H \propto \Delta V_0^a$. Then for each galaxy

$$\begin{aligned} -\Delta V_G &= V_c - V_H - V_{\text{pec}} \\ &= V_c - 10^{1/2((HM)+a \log \Delta V_0 - \log F_H)} - V_{\text{pec}} \end{aligned} \quad (2)$$

We can use $\langle HM \rangle$, the average value of HM , and assume that

V_{pec} is randomly distributed so that its expected value is zero to obtain a set of linear equations of the form

$$\begin{aligned} \Delta V_{Gi} &= l_i X + m_i Y + n_i Z \\ &= V_c - 10^{1/2((HM)+a \log \Delta V_0 - \log F_H)} \end{aligned} \quad (3)$$

which relates ΔV_{Gi} to observed quantities where X , Y and Z are the components of the Local Group motion and l_i , m_i and n_i are the direction cosines of the i th galaxy. A least-squares solution of this over-determined set of equations gives an estimate of the vector V_G .

A second and more rigorous statistical approach is to treat this as a well-defined non-linear minimization problem. We take the relation between M_H and ΔV_0 to be of the form

$$M_H = b \Delta V_0^a \quad (4)$$

M_H is related to the observed integrated flux density by $M_H = \text{const } F_H \cdot V_H^2$ where V_H is the Hubble velocity representing the true distance as defined above. Accordingly an estimate of V_H is

$$V_H = C \Delta V_0^{a/2} F_H^{1/2} \quad (5)$$

In the analysis V_{pec} is assumed to be normally distributed with a mean value of zero. The quantity to be minimized in determining the Local Group motion (X , Y , Z) is

$$\begin{aligned} \sum V_{\text{pec}i}^2 &= \sum [V_{ci} - (V_{Hi} + l_i X + m_i Y + n_i Z)]^2 \\ &= \sum [V_{ci} - C \Delta V_{0i}^{a/2} F_{Hi}^{1/2} - (l_i X + m_i Y + n_i Z)]^2 \end{aligned} \quad (6)$$

The minimization is performed with respect to a , C , X , Y and Z . The solutions obtained by this more rigorous method agree well with those using the first method thereby indicating that the details of the analysis procedures are not too important for this set of data.

The main analysis described here is based on 84 Sbc galaxies with an inclination $i > 30^{\circ}$ ($i = 0^{\circ}$ for face-on galaxies) and $V_c > 1,000 \text{ km s}^{-1}$. Figure 1a gives the velocity spread of the galaxies used in this analysis; the median velocity is 1,800 km s^{-1} . For galaxies more face-on than $i = 30^{\circ}$ the inclination, and hence $\Delta V_{1/2} \csc i$, becomes more uncertain. By making solutions for the Local Group motion with different cutoff values of i , we find that they are not strongly dependent on the cutoff value.

The motion found for the Local Group is $383 \pm 50 \text{ km s}^{-1}$ towards $l = 256 \pm 18^{\circ}$, $b = +52 \pm 16^{\circ}$ (supergalactic coordinates $L = 110 \pm 16^{\circ}$, $B = -27 \pm 14^{\circ}$). This solution corresponds

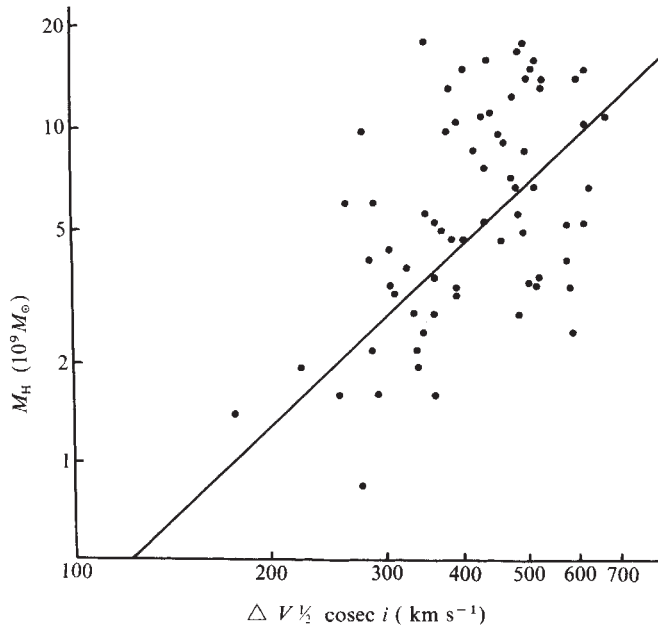


Fig. 2 The correlation between the neutral hydrogen mass M_H of Sbc ($T=4$) galaxies and their profile widths $\Delta V_{1/2} \csc i$. $\Delta V_{1/2}$ is the width at half-power and i is the inclination of the galaxy. M_H is the neutral hydrogen luminosity indicator used in the solution for the Local Group motion. $M_H = F_H D^2$ where F_H is the integrated flux density and D is the distance in Mpc.

to a solar velocity of $314 \pm 50 \text{ km s}^{-1}$ towards $\text{RA} = 12.9 \text{ h} \pm 0.7 \text{ h}$, $\text{dec.} = 10^\circ \pm 12^\circ$ (galactic coordinates $l = 309^\circ \pm 35^\circ$, $b = 73^\circ \pm 12^\circ$). The standard motion of the Sun relative to the Local Group of galaxies of 300 km s^{-1} towards $l = 90^\circ$, $b = 0^\circ$ is used throughout.

Figure 2 shows the correlation between the neutral hydrogen luminosity of each galaxy, expressed as the total mass of neutral hydrogen, $M_H = F_H D^2$, plotted as a function of the observed profile width, ΔV_0 , corrected for inclination. This correlation was obtained from the fitting procedures described above and yields a value of $a = 1.85 \pm 0.25$. Essentially each galaxy in Fig. 2 is placed at its 'true' redshift distance D based on the observed redshift and the derived value of the solar motion.

An indication of the validity of the analytical solution for the Local Group motion obtained here can be obtained by simply plotting as a function of supergalactic longitude L the difference between the observed redshift and the luminosity distance expressed as a redshift. Because the Local Group motion is in a direction close to the supergalactic plane this plot should show a sinusoidal variation of velocity difference with longitude. The amplitude plotted in Fig. 3 is 436 km s^{-1} directed towards $L = 119^\circ$. The plot includes only the galaxies with $1,000 < V_c < 3,200 \text{ km s}^{-1}$ and supergalactic latitude $|B| < 50^\circ$. The full solution in Table 1, of course, includes all the galaxies in the sample with $V_c > 1,000 \text{ km s}^{-1}$ and covers the whole sky.

It is relevant at this stage to comment on the possible effects on the solutions for the Local Group motion of Malmquist bias^{6,16} in which the more distant galaxies in a magnitude limited sample tend to have higher intrinsic luminosity (or H I mass in our case). A plot of M_H against redshift shows only a weak Malmquist bias, largely because the galaxies in both the Sbc and Sc samples have a restricted luminosity class, corresponding to a restricted range of luminosity or H I mass. This latter property is illustrated by the small ($\sim$ a factor of 2) range of ΔV_0 seen in Fig. 2. More important, the effects of any residual bias in the data are removed if there is a uniform distribution of galaxies in redshift around the sky; our galaxies satisfy this criterion. Further, the effects of Malmquist bias are also largely eliminated by the use of a third variable, ΔV_0 , in our case. We

are confident that the effects of Malmquist bias do not seriously affect our solution for the Local Group motion. This is confirmed by the fact that we obtain similar solutions for the Sbc and Sc galaxies and by using galaxies at different redshifts. These important matters will be discussed in more detail elsewhere.

The Virgo cluster of galaxies probably had a significant influence on the motion of the galaxies lying in the neighbourhood of the cluster; indeed we shall argue below that the Virgo cluster is responsible for the major part of the Local Group motion. Two solutions have been made using different criteria for excluding galaxies influenced by the Virgo cluster. In the first, members of the Virgo cluster are excluded; the Local Group is found to move at 434 km s^{-1} towards $l = 262^\circ$, $b = 52^\circ$. In the second, those galaxies with a line of sight component of Virgo-centric infall velocity greater than our own are eliminated; the Local Group is then found to move at 436 km s^{-1} towards $l = 264^\circ$, $b = 45^\circ$. We consider this to be our best estimate of the Local Group motion based on the H I observations of Sbc galaxies.

A value for the Local Group motion has also been derived from the optical data for the Sbc galaxies using an approach of minimum sophistication. The blue luminosities in RC2 were used as standard candles using the velocity width ΔV_0 as a third variable as for the H I analysis. This solution has lower precision than that obtained from the H I data; the solution is $397 \pm 90 \text{ km s}^{-1}$ towards $l = 216^\circ \pm 20^\circ$, $b = 27^\circ \pm 15^\circ$, a value similar to the H I result.

A completely independent data set is available for an analysis identical to that performed on our Sbc H I data. Rubin *et al.*⁶ give H I parameters and plot H I spectra for a substantial fraction of their list of Sc galaxies. We have used all their galaxies with $1,000 < V_c < 5,000 \text{ km s}^{-1}$ and supplemented them with southern Sc galaxies from van Woerden *et al.* (in preparation) to provide a full sky coverage. Figure 1b shows the velocity distribution of this set of galaxies. The 53 Sc galaxies in this combined data set give a solution for the Local Group motion of 580 km s^{-1} towards $l = 245^\circ$, $b = 35^\circ$. This motion is consistent with the result obtained for the Sbc galaxies.

The results obtained from the present study are compared in Table 1 and Fig. 4 with the motion of the Sun and the Local Group relative to the microwave background derived on the assumption that the dipole anisotropy is entirely the result of solar motion. For a microwave background temperature of 2.7 K (ref. 17), the mean of the recent dipole determinations by Boughn *et al.*² and by Gorenstein and Smoot¹ implies a solar motion of $410 \pm 45 \text{ km s}^{-1}$ towards $\text{RA} = 11.4 \text{ h} \pm 0.3 \text{ h}$ and $\text{dec.} = 3^\circ \pm 15^\circ$ ($l = 259^\circ$, $b = 58^\circ$). The corresponding motion of the Local Group is $620 \pm 50 \text{ km s}^{-1}$ towards $l = 265^\circ \pm 9^\circ$, $b = 35^\circ \pm 7^\circ$ ($L = 128^\circ$, $B = -36^\circ$). If the microwave background temperature is 3.0 K (ref. 18) instead of 2.7 K the solar velocity is reduced to 370 km s^{-1} towards $l = 259^\circ$, $b = 58^\circ$ and the Local

Table 2 Component of the motion of the Local Group towards the Virgo cluster*

Method	Velocity (km s^{-1})	Ref.
(1) H I measurements of nearby Sbc and Sc galaxies with $1,000 < V_c < 5,000 \text{ km s}^{-1}$	410 ± 55	This paper
(2) 2.7 K cosmic microwave background, dipole anisotropy	$450 \pm 50^\dagger$	1, 2
(3) Four nearby clusters in IR with Fisher-Tully correction	480 ± 75	23
(4) E galaxies in and around the Virgo cluster	470 ± 75	24
(5) Mean value of methods 1, 2, 3, 4	447 ± 40	

* The Virgo cluster is assumed to be centred at $l = 280.4^\circ$, $b = 74.6^\circ$ ($L = 102.2^\circ$, $B = -3.0^\circ$). See ref. 26.

[†] Value for a background temperature of 2.7 K ; the value is 430 km s^{-1} for a background temperature of 3.0 K .

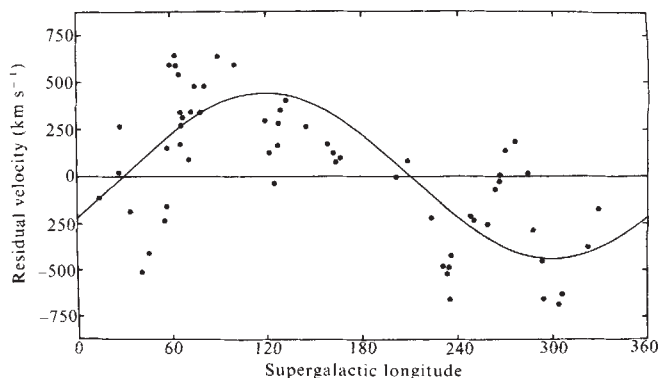


Fig. 3 The velocity residual between the luminosity distance of Sbc ($T=4$) galaxies estimated as a redshift and the observed redshift plotted as a function of supergalactic longitude, L . Only galaxies with $1,000 < V_c < 3,500 \text{ km s}^{-1}$ and within 50° of the supergalactic plane are included. The velocity residual shows a sinusoidal variation with L which is a consequence of the Local Group motion towards $L \sim 120^\circ$. The sinewave plotted is that expected for a velocity of 436 km s^{-1} towards $L = 119^\circ$, the solution found from a full analysis of all the Sbc galaxies; it is not a fit to plotted points alone—see text.

Group velocity reduces to 580 km s^{-1} towards $l = 266^\circ$, $b = 32^\circ$. The magnitude and direction of the solar motion determined from the H I observations of Sbc and Sc galaxies agree with the microwave dipole values to within two standard deviations (s.d.) of the combined errors.

The present results emphasize the disagreement between the Rubin *et al.* optical result^{6,7} and that obtained from the microwave background measurements. The agreement found here between the determination based on H I data for the Rubin *et al.* galaxies and the other results in Table 1 confirms that the optical data base or the reduction procedures of Rubin *et al.* are in error^{11,19}. Note that de Vaucouleurs and colleagues^{12–14} find a solution for the solar motion calculated relative to the backdrop of spiral galaxies which is similar to our optical result. However, their solution is displaced by 4 s.d. from the statistically more accurate H I solutions which we believe are representative of the backdrop of galaxies since the H I sample is relatively free of contamination by galaxies whose velocities may be strongly affected by the Virgo cluster. The de Vaucouleurs galaxy samples on the other hand include $\sim 30\%$ with heliocentric velocities $< 1,000 \text{ km s}^{-1}$ and a substantial number ($\sim 10\%$) of Virgo cluster members.

Consequences of Local Group motion

The close agreement between the velocity of the Local Group calculated relative to the backdrop of nearby galaxies and that derived from the 3 K cosmic microwave background dipole anisotropy argues strongly that the anisotropy is produced by the Local Group motion. This agreement is all the more significant as the velocity is directed to within 35° of the Virgo cluster ($l = 280^\circ$, $b = 75^\circ$, $L = 102^\circ$, $B = -3^\circ$) which is likely to have a major influence on the motion of the Local Group. The mean value of the motion derived from the two recent and independent microwave background results and from the two H I results is $546 \pm 70 \text{ km s}^{-1}$ towards $l = 261^\circ \pm 9^\circ$, $b = 39^\circ \pm 7^\circ$ ($L = 122^\circ \pm 9^\circ$, $B = -35^\circ \pm 7^\circ$). We consider that this is the best estimate of the Local Group motion available.

It is still possible that a fraction of the 3 K dipole anisotropy is intrinsic. An estimate of this intrinsic anisotropy can be derived from the velocity difference between the Local Group motion calculated relative to the backdrop of galaxies and that relative to the 3 K microwave background. The velocity difference is $130 \pm 70 \text{ km s}^{-1}$ for an assumed microwave back-

ground temperature of 2.7 K or $90 \pm 65 \text{ km s}^{-1}$ for 3.0 K . The corresponding difference amplitudes of the dipole component are 1.3 ± 0.7 and $0.9 \pm 0.7 \text{ mK}$ or 26–37% of the observed dipole anisotropy of $\sim 3.5 \text{ mK}$.

Our present results therefore rule out the possibility that most of the cosmic background dipole anisotropy is intrinsic^{4,5,20} and is due to an anisotropic cosmology or a large scale anisotropic matter distribution in the Universe. The relatively small velocity difference between the microwave background and the backdrop of galaxies implies much lower amplitude long-range galaxy correlations than suggested by Clutton-Brock and Peebles⁸. Further, our results raise some doubt about whether the tentative detections of some quadrupole components in the microwave background can be interpreted in terms of the anisotropic models of Silk and Wilson⁴ or of Peebles⁵. Silk and Wilson predict quadrupole components which are 3×10^{-1} to 10^{-3} of the dipole component depending on the index of the power-law fluctuation spectrum; Peebles predicts a ratio of 3×10^{-1} . Our limit on the intrinsic dipole anisotropy apparently indicates that some reported quadrupole components are comparable with or greater than the intrinsic dipole component in disagreement with the above predictions.

Component of velocity towards the Virgo cluster

An evaluation of the component of the Local Group motion towards the Virgo cluster provides a means^{21,22} of estimating the density of the Universe. Several estimates of this Virgo-centric motion are already available; these will be compared and combined with the present determination of the Virgo-centric motion.

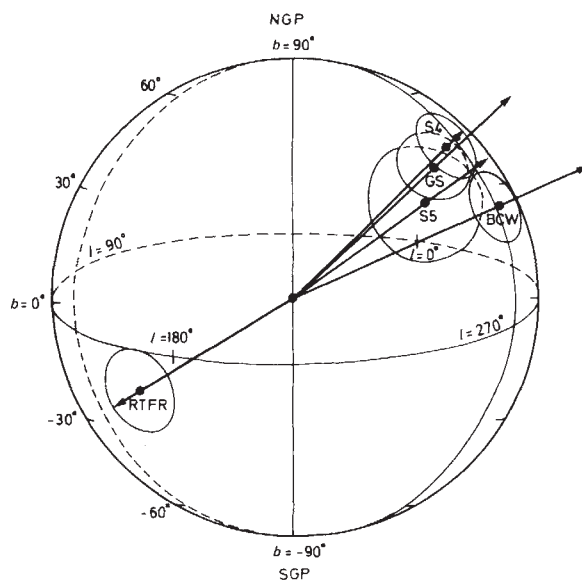


Fig. 4 Various solutions for the motion of the Local Group of galaxies plotted as vectors in galactic coordinates. Each solution is shown as a velocity with the appropriate length and with an error in angle described by a small circle on the surface of the celestial sphere. The vectors are coded as follows; S4, the Sbc ($T=4$) H I data of the present study; S5, the Sc ($T=5$) H I of the present study; GS, the microwave background dipole anisotropy from Gorenstein and Smoot; BCW, the microwave anisotropy from Boughn, Cheng and Wilkinson; RTFR, the Sbc galaxy optical data of Rubin *et al.* The agreement between the microwave background results and the H I results are evident. The Rubin *et al.* results are clearly discordant. The centre of the Virgo cluster lies at $l = 280^\circ$, $b = 75^\circ$ (from ref. 26).

The H I solution for the Sbc and Sc galaxies and for the microwave background give a Virgo-centric component of Local Group motion of $\sim 450 \text{ km s}^{-1}$ (Table 2). Another estimate of high precision ($> 5 \text{ s.d.}$) comes from work of Aaronson *et al.*²³ who used IR fluxes of individual galaxies corrected by the Fisher–Tully relation in four nearby clusters at redshifts in the range $4,000\text{--}6,000 \text{ km s}^{-1}$. Tonry and Davis²⁴ determined the Virgo-centric velocity from a study of E galaxies in and around the Virgo cluster. There is remarkably good agreement between these four high accuracy results.

Other estimates of the Virgo-centric motion^{9,19,25} are available with lower quoted accuracy. Some of these refer to nearby galaxy samples which may themselves be influenced by the gravitational potential of the Virgo cluster. We consider that these results are not in serious agreement with the adopted Virgo-centric infall velocity of $447 \pm 50 \text{ km s}^{-1}$.

Finally, we comment on the component of Local Group motion which is orthogonal to the direction of the Virgo cluster. The best estimate of the motion of the Local Group of 546 km s^{-1} towards $l = 261^\circ \pm 9^\circ$, $b = 39^\circ \pm 7^\circ$ (Table 1) is in a direction lying outside the errors of measurement of direction given by Sandage and Tammann²⁶ for the centre of luminosity of the Virgo cluster at $l = 280^\circ$, $b = 75^\circ$. Assuming that this is also the mass centre of the Virgo cluster, the difference in direction may be interpreted as the result of a component of motion in a direction orthogonal to the line of sight to the Virgo cluster. This motion is $320 \pm 70 \text{ km s}^{-1}$ towards $l \sim 254^\circ$, $b \sim -14^\circ$. Taken at its face value this motion is a 4.5 s.d. result for a Hubble flow departure in a velocity of $2,500 \text{ km s}^{-1}$, the mean of the Sbc and Sc galaxies. Such a motion provides a deviation $\Delta H_0/H_0$ in the Hubble flow of $320/2,500 = 0.13$ and is not at variance with the known departures²⁷ from the mean Hubble flow. However, the major component (300 km s^{-1}) of this orthogonal motion is directed towards the supergalactic plane and may be the result of collapse onto the flattened Virgo supercluster mass distribution²⁸. The resultant motion in the Supergalactic plane perpendicular to the Virgo direction would be 100 km s^{-1} leading to $\Delta H_0/H_0 = 0.04$ for the random motion of the Local Group.

Virgo-centric motion and estimates of Ω and q_0

The evidence from the dipole anisotropy in the microwave background and the motion of the Sun relative to the backdrop of nearby galaxies strongly suggests that the Local Group is falling towards the centre of the Virgo cluster at a velocity of $447 \pm 40 \text{ km s}^{-1}$. Such a departure from the Hubble flow of the Universe can be used to derive a local value for the deceleration parameter q_0 and for the ratio, $\Omega (= 2q_0$ for Friedmann cosmologies), of the local density to the closure density of the Universe^{19–22,24,29,30}. The calculation of q_0 and Ω requires an estimate of the density enhancement δ in the Virgo cluster relative to that in the smoothed-out Universe. Yahil *et al.*^{29,30} find $\delta = 3.5 \pm 0.4$ while Davis *et al.*³¹ find a value of 2.0 ± 0.3 from a different sample of galaxies. A detailed treatment^{22,31} of the peculiar velocity v_p produced by a spherically symmetric

density distribution gives, from a linear perturbation analysis,

$$\Omega = \left(\frac{3v_p}{\delta(v_p + v_{VC})} \right)^{1.67} \quad (7)$$

where v_{VC} is the Virgo cluster velocity relative to the Local Group and is taken as $1,060 \text{ km s}^{-1}$. This yields a value of $\Omega = 0.10\text{--}0.26$ for $v_p = 447 \text{ km s}^{-1}$ and δ in the range 2–3.5. Using the results of a more exact non-linear calculation³¹ we find $\Omega = 0.18 \pm 0.04$ to 0.39 ± 0.11 for this range of δ . Clearly the high value of the peculiar Virgo-centric velocity found in the present results and in the microwave background results implies a value of local density approaching the closure density, with $\Omega = 2q_0 = 0.15\text{--}0.50$ where the higher values correspond to the lower value of density enhancement ($\delta = 2$) in the Virgo cluster. Although this high value of Ω has been found for the local region of the Universe, essentially the local supercluster, there is no reason to believe that it is in any way different from the Universe at large. The infall velocity from which it is derived is the result of an acceleration acting over the age of the Universe.

Conclusion

The present observations demonstrate the use of a method of obtaining the statistical distances of spiral galaxies using H I data alone, namely H I fluxes and profile velocity widths. This method seems to be comparable in accuracy with related optical methods and has the advantage of requiring fewer correction factors. When applied to the determination of the motion of the Local Group of galaxies relative to the backdrop of nearby Sbc galaxies the method gives a velocity of $436 \pm 55 \text{ km s}^{-1}$ towards $l = 264^\circ \pm 18^\circ$, $b = 45^\circ \pm 12^\circ$, in agreement with the velocity implied by the dipole anisotropy in the 3 K cosmic microwave background.

The agreement between the motion of the Local Group determined relative to the nearby galaxies and that inferred from the microwave background is of fundamental significance. It implies that the major part at least of the dipole anisotropy is extrinsic, due to the Local Group motion, and is not intrinsic as implied in some cosmologies. The maximum intrinsic dipole anisotropy allowed by the combined data is $\sim 1 \text{ mK}$, indicating that $\Delta T/T \leq 3 \times 10^{-4}$ for any intrinsic dipole anisotropy in the temperature distribution of the early ($z \sim 1,500$) Universe. The claimed quadrupole anisotropy² cannot then be the result of the processes envisaged by Silk and Wilson⁴ and by Peebles⁵.

The Virgo-centric velocity of the Local Group of galaxies is now well-determined from several independent methods. The inferred local values of q_0 and Ω depend on the assumed model of the mass density distribution in the Virgo cluster. In any case, high values of Ω (0.15–0.50) are implied by the observed Virgo-centric velocity. Such values, if applicable to the Universe as a whole, will produce a significant deceleration of the expansion of the Universe.

We thank Mr S. C. Johnson for discussions regarding the H I data base and Dr A. N. Lasenby regarding statistical techniques, also Professor H. van Woerden for supplying in advance of publication the southern H I data.

Received 21 September 1981; accepted 31 March 1982.

- Gorenstein, M. V. & Smoot, G. F. *Astrophys. J.* **244**, 361–381 (1981).
- Boughn, S. P., Cheng, E. S. & Wilkinson, D. T. *Astrophys. J. Lett.* **243**, L113–L117 (1981).
- Fabbri, R., Guidi, I., Melchiorri, F. & Natale, V. *Phys. Rev. Lett.* **44**, 1563–1566 (1980).
- Silk, J. & Wilson, M. L. *Astrophys. J. Lett.* **244**, L37–L41 (1981).
- Peebles, P. J. E. *Astrophys. J. Lett.* **243**, L119–L122 (1981).
- Rubin, V. C., Ford, W. K., Thonnard, N., Roberts, M. S. & Graham, J. A. *Astr. J.* **81**, 687–718 (1976).
- Rubin, V. C., Thonnard, N., Ford, W. K. & Roberts, M. S. *Astr. J.* **81**, 719–737 (1976).
- Clutton-Brock, M. & Peebles, P. J. E., *Astr. J.* **86**, 1115–1119 (1981).
- Tammann, G. A., Sandage, A. & Yahil, A. *Phys. Scr.* **21**, 630–634 (1980).
- Fall, S. M. & Jones, B. J. T. *Nature* **262**, 457–460 (1976).
- Schechter, P. L. *Astr. J.* **82**, 569–576 (1977).
- de Vaucouleurs, G. & Bollinger, G. *Astrophys. J.* **233**, 433–452 (1979).
- de Vaucouleurs, G. & Peters, W. L. *Astrophys. J.* **248**, 395–407 (1981).
- de Vaucouleurs, G., Peters, W. L., Bottinelli, L., Gouguenheim, L. & Paturel, G. *Astrophys. J.* **248**, 408–422 (1981).
- de Vaucouleurs, G., de Vaucouleurs, A. & Corwin, H. *Second Reference Catalogue of Bright Galaxies* (University of Texas, Austin 1976).
- Sandage, A., Tammann, G. A. & Yahil, A. *Astrophys. J.* **232**, 352–364 (1979).
- Wilkinson, D. T. *Phys. Scr.* **21**, 606–609 (1980).
- Richards, P. L. *Phys. Scr.* **21**, 610–613 (1980).
- Yahil, A. *10th Texas Symp.*, (New York Academy of Science, in the press).
- Wilson, M. L. & Silk, J. *Astrophys. J.* **243**, 14–25 (1981).
- Silk, J. *Astrophys. J.* **193**, 525–527 (1974).
- Peebles, P. J. E. *Astrophys. J.* **205**, 318–328 (1976).
- Aaronson, M. *et al. Astrophys. J.* **239**, 12–37 (1980).
- Tonry, J. L. & Davis, M. *Astrophys. J.* **246**, 680–695 (1981).
- Schechter, P. L. *Astr. J.* **85**, 801–811 (1980).
- Sandage, A. & Tammann, G. A. *Astrophys. J. Lett.* **207**, L1–L4 (1976).
- Sandage, A. *Astrophys. J.* **202**, 563–582 (1975).
- White, S. D. M. & Silk, J. *Astrophys. J.* **231**, 1–9 (1979).
- Yahil, A., Sandage, A. & Tammann, G. A. *Phys. Scr.* **21**, 635–639 (1980).
- Yahil, A., Sandage, A. & Tammann, G. A. *Astrophys. J.* **242**, 448–468 (1980).
- Davis, M., Tonry, J., Huchra, J. & Latham, D. W. *Astrophys. J. Lett.* **238**, L113–L116 (1980).

Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences

David Penny*, L. R. Foulds† & M. D. Hendy‡

* Department of Botany and Zoology, and ‡ Department of Mathematics and Statistics, Massey University, Palmerston North, New Zealand
† Operations Research, University of Canterbury, Christchurch, New Zealand

The theory of evolution predicts that similar phylogenetic trees should be obtained from different sets of character data. We have tested this prediction using sequence data for 5 proteins from 11 species. Our results are consistent with the theory of evolution.

THE theory of evolution continues to be a focus for nearly all biological thought. Nevertheless, there have been doubts about the philosophical status of the theory, particularly on the extent to which it can be tested or falsified. The best known of these doubts have been expressed by the leading philosopher on scientific method, Karl Popper¹, who concluded that "darwinism is not a testable scientific theory, but a 'metaphysical research program'—a possible framework for testable scientific theories". Popper did not in any way reject evolution. He pointed out (ref. 1, p. 169) that "no serious competitor has come forward" and commented on "the strange similarity between my theory of the growth of knowledge and darwinism". In particular, he suggests that only one prediction is possible from darwinism: "Gradualism is thus, from a logical point of view, the central prediction of the theory. (It seems to me that it is the only prediction.)" (ref. 1, p. 172).

Popper has recently modified these criticisms², pointing out that criticisms by some authors were inconsistent, such as that natural selection is a tautology, and that it explains too much. A tautology explains nothing, so cannot simultaneously explain "too much". In addition, the suggestion was made that the existence of an evolutionary tree was falsifiable, but no reasoning was given for this new opinion.

These criticisms have aroused considerable interest^{3–5}. Most of the discussion has, however, been qualitative, so it would be useful to be able to make quantitative tests. This is the purpose of the present article.

Popper makes the important distinction between the existence of an evolutionary tree (an evolutionary history) and the mechanism of evolution put forward to explain the processes that produced that history. Here we present a programme, applying graph theory^{6,7}, by which it is theoretically possible to refute the existence of an evolutionary tree.

Another prediction from evolution

It has been long been considered that protein sequence data contain evolutionary information⁸. In particular, the tree of minimal length (minimum number of mutations, maximum

parsimony) makes no assumptions about the mechanism of evolution, and has been widely used as a model for evolutionary relationships^{9–11}. Given any comparative data, irrespective of origin, one can construct trees of an evolutionary form, and hence a minimal tree must exist. So in general, finding a minimal tree for protein sequence data is not in itself independent evidence for the existence of an evolutionary tree.

However, another prediction can be made if there has been an evolutionary tree and if the maximum parsimony model is a good predictor of that tree. The prediction is that minimal trees with the same taxa should be similar, or 'congruent'¹², when constructed from different protein sequences. We need a measure of tree similarity having biological significance so that the values can be compared with the hypothesis that the trees are randomly selected. Such a measure is described below.

Our strategy is to take different protein sequences for a common set of taxa, find all the minimal (and near minimal) evolutionary trees and then compare them. Should the probability be high that these trees are unrelated, this would indicate that the protein sequences do not contain similar evolutionary information, and hence would contradict the existence of an evolutionary tree for those taxa.

We conclude that the existence of an evolutionary tree for these taxa is a falsifiable hypothesis, and therefore meets Popper's criteria for scientific theories. In addition, our methods allow us to identify a consensus tree which incorporates the most common features of all the near minimal trees.

Finding minimal evolutionary trees

It is easily shown that any minimal evolutionary tree can be represented by a binary tree. Using the double factorial notation (!) there are $(2n-5)!! = 1 \times 3 \times 5 \dots \times (2n-5)$ unrooted binary trees spanning n sequences^{13,14}. To determine the trees of maximum parsimony (minimal length), one must potentially consider a vast number of different tree topologies. But for all but a small number of sequences ($n \leq 8$), present day computers cannot consider them all¹⁰.

Table 1 The frequency $p(m, n)$ of occurrence of $d(T_i, T_j) = m$ for binary trees spanning n taxa $4 \leq n \leq 11$

	$m=0$	2	4	6	8	10	12	14	16	$E(n)$
$n=4$	3.3×10^{-1}	6.7×10^{-1}								1.34
5	6.7×10^{-2}	2.7×10^{-1}	6.7×10^{-1}							3.20
6	9.5×10^{-3}	5.7×10^{-2}	2.4×10^{-1}	7.0×10^{-1}						5.24
7	1.1×10^{-3}	8.5×10^{-3}	4.6×10^{-2}	2.2×10^{-1}	7.3×10^{-1}					7.32
8	9.6×10^{-5}	9.6×10^{-4}	6.5×10^{-3}	3.8×10^{-2}	2.0×10^{-1}	7.5×10^{-1}				9.40
9	7.4×10^{-6}	8.9×10^{-5}	7.0×10^{-4}	4.8×10^{-3}	3.1×10^{-2}	1.9×10^{-1}	7.7×10^{-1}			11.46
10	4.9×10^{-7}	6.9×10^{-6}	6.2×10^{-5}	4.8×10^{-4}	3.6×10^{-3}	2.7×10^{-2}	1.8×10^{-1}	7.9×10^{-1}		13.52
11	2.9×10^{-8}	4.6×10^{-7}	4.7×10^{-6}	4.0×10^{-5}	3.3×10^{-4}	2.8×10^{-3}	2.3×10^{-2}	1.7×10^{-1}	8.1×10^{-1}	15.55

$E(n) = \sum mp(m, n)$ is the weighted mean or 'expected value' of $d(T_i, T_j)$ for randomly selected binary trees T_i, T_j spanning n taxa.

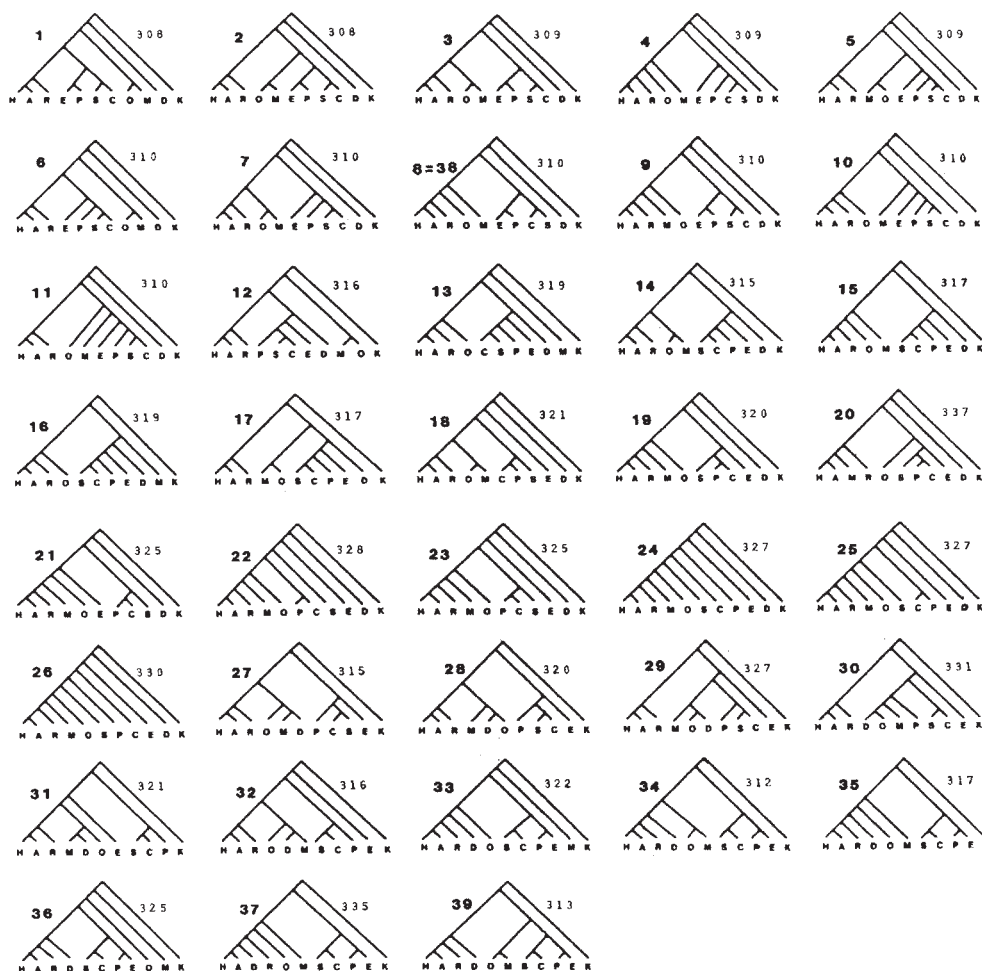


Fig. 1 The 39 minimal and near minimal trees spanning the 11 taxa given in Table 2. T1–T11 are generated from the complete sequences, T12–T17 from cytochrome *c*, T18 from fibrinopeptide A, T19–T26 from fibrinopeptide B, T27–T32 from haemoglobin A, and T33–T39 from haemoglobin B. The trees are arranged with their sequences in increasing lengths (see Table 3). The total number of mutations of each tree for the combined sequences is shown. The lengths of individual links are not shown, only the branching sequence is indicated. See Table 2 legend for definition of letter abbreviations.

We have shown that the problem of finding minimal length trees is an example of the Steiner problem in graphs^{15,16} which makes no assumptions about the existence of an evolutionary history. Two recent techniques developed by the authors (refs 15–17 and D.P., in preparation) have enabled us to solve this for some larger sets of sequences (in one case a tree with 25 species has been proved minimal). These techniques rely on the inherent structural content of the data and are found not to work well on random data. This is an indication that there may be considerable tree-like structure within the sequence data. However, these observations are not easily quantifiable and the inability to use such methods on some data sets of this size ($8 < n < 25$) could not be taken as evidence for or against the existence of an evolutionary tree.

Here we have used a recently developed (ref. 17 and D.P., in preparation) 'branch and bound' method⁷ which, for the data in Table 2, has found all minimal and near minimal trees.

Comparison of trees

Several methods have been proposed¹⁸ that measure some features of the difference between a pair of trees, but it is important that any method chosen has biological relevance. Apart from satisfying a historical curiosity, a major application of an evolutionary tree is to assist in the classification of the taxa into groups at different levels. The set of taxa that are descendants of a given common ancestor will form a group of related taxa. In a strictly bifurcating (binary) tree spanning n taxa, there will be $n-2$ such non-trivial classes of the taxa. As in Robinson and Foulds¹⁹, we measure the difference $d(T_1, T_2)$ of two trees T_1 and T_2 , as the number of classes which are derived from T_1 or T_2 , but not both. It is easily shown that this measure forms a metric on the space of all trees spanning these n taxa.

Given a non-directed tree T (one in which the root has not been specified), the removal of any internal link will partition the taxa into two subsets, one of which (depending on orientation) will be a group of related taxa. Waterman and Smith²⁰ have shown that the set of all such partitions uniquely defines T . Thus, $d(T_1, T_2)$ also represents the number of partitions of the taxa formed by deleting internal links, which differ in the two trees. This analysis is independent of whether or not the tree is a rooted tree.

Probabilities and tree comparisons

We can, for any specific tree T , determine the value of $d(T, T')$ for each of the $(2n-5)!!$ trees T' . Then the proportion of trees with $d(T, T') = m$ will represent the probability that a randomly selected tree T' is distance m from T .

For binary trees the number of internal links is $n-3$, so each tree defines $n-3$ partitions. If T_i, T_j have m partitions in common, $d(T_i, T_j) = 2(n-3-m)$ which is even, and can range from 0 to $2n-6$. The value $d(T_i, T_j) = 0$ can occur only for $i=j$, so the frequency of $d(T_i, T_j) = 0$ is $1/N$, where $N = (2n-5)!!$. If $d(T_i, T_j) = 2$ then there is only one link at which the two trees disagree. If we delete one internal link of T_i , there are two alternative ways of rejoining it to give values $d(T_i, T_j) = 2$. This could occur at any of the $n-3$ internal links, so we find $d(T_i, T_j) = 2$ with frequency $(2n-6)/N$.

The values of $d(T_i, T_j)$ over all pairs of binary trees spanning n species for $4 \leq n \leq 11$ have been determined using recursive generating functions (D.P. and M.D.H., in preparation). These results are summarized in Table 1.

We can, for a given value of m , use these frequencies to estimate the probability of randomly selecting two trees T_i, T_j with $d(T_i, T_j) = m$. Referring to the case $n=11$, we would, for example, expect $d(T_i, T_j) = 4$ to occur with probability

Table 2 Nucleotide sequences derived from the 5 polypeptides for 11 taxa

Taxon	Haemoglobin A	Haemoglobin B
R	CUCGGGGGGAUUACGGAAACUAGGAAAGGCAGCUGCAGAG	GCCACA AACACCCCCAUCUCCCGGCGGGACCAACCACAAGCUGACUAACAGCA
S	GAGGGGGGCAUUACGGGGCAAAAGCCGGAGACUGUAACAGA	AACAGGCACGCUCCGAUCACGCGACGAUACAAAACGGCAGCGUGAUCAGUCAU
E	GACGGGGGGAUUGCGCAAACUAGCCGGGGACGUUAACGAG	GCGUGGC UUGCCAGCUC AACAGGCCAUGAGCACCAGCGGCGUGAUACGACCA
K	GGGGAGGGGACAACACAAGCAAAACGGGAAAGCUGAAGGC	GCCAGAAACAGCAAAACCAAGAGCCGUGGCAAAAAAAAAAAAAUAAUAGAAGC
M	GUCAGGAGGAUCACGGAAGAACCCCGGGGCAUCAGCCGAG	GCAUGGCUCGGCCCGAUCUCGCGGGACGACCAACCAUCGAAAUACCACGCGCU
O	CACAGAAGGAUUGAACAAAAAGCCGGGACAAAUUCAGAA	GCGAAUCACGCCAGCAUCUCGCAACGCGAACAGACAAAAGGCAUACUAACAGCA
D	CACAAAAGGCCAACGCAAAACACCCGGGGCAUGAACAGGC	GCCAGUCUCGGCCCAUCACCCAGGAUAUACCAAAAAAAGCUGACUAACAGCA
P	GGCGGGGCGAUUGAGCGAAAAAGCAGGGGCAUCUAACAAA	GCCUGGC UUGGCCCGAUCAAGCGGCGAUACAACAACAAAGGAGUGAUCCGUCCA
H	CACGGGGGCAUUACGGAAAAACCGAAAGGCAGCUGCAGAG	GCCACUCACGCCCCGAUCACCCGGCGGGACCGCCCAACGGCUGACUAACCGCA
C	GGGGGGGGCAUUACGGGCAAAAACCGGGGAAUCUAGCGAA	AACAGGCACGGUACGACAAAGCAACGAUACACAACGCAGCGUGACCAGUCAU
A	CACGGGGGCGUUAACGGAAAAACCGAAAGGCAGCUGCAGAG	GCCACUCACGCCCCGAUCACCCGGCGGGACCGCCCAACAGCUGACUAACCGCA

	Fibrinopeptides A and B	Cytochrome c	
R	GAGGGAGGACCCG	AGAGAGAAGCCUAG	AAUUCAUCCCUAC
S	GCGUGGGGAACCG	AGGGCAACUCCUGA	GCAGCUUACACGA
E	CAGAGAGGAACAG	AGGACAAGUACUGA	GCAGCUAACACCA
K	CAAAGAGACAACG	AGGGUAAAGAGGGA	GCAGCUAUCCCGC
M	GAGGACAGAAAAC	AGCAAGUGAAUUG	GCAGGUUACCCGC
O	GCGGGAACAAAC	AGCGAGAGUUUCGA	GCAGGUUACCCAC
D	AUAAGAGGAAACG	AGGAUGUAGACGGA	GCAGCUUACACGC
P	GGCAAAGGAACCG	AGGCCAAGUCAGGA	GCAGCUUACACGA
H	GUGGGAGGACCCG	UAAGAGAGGUUUAG	AAUUCAUCUCUAC
C	GCCCAGGGACCCG	AGGCCAAGUGGUGG	GCAGCUUACACGA
A	GUGGGAGGACCCG	UAAGAGAGGUUUAG	AAUUCAUCUCUAC

R = Rhesus monkey (*Macaca mulatta*); S = sheep (*Ovis ammon*); E = horse (*Equus caballus*); K = kangaroo (*Macropus congruus* or *Macropus giganteus*); M = mouse (*Mus musculus* or *Rattus rattus*); O = rabbit (*Oryctolagus cuniculus*); D = dog (*Canis familiaris*); P = pig (*Sus scrofa*); H = human (*Homo sapiens*); C = cow (*Bos primigenius*); A = ape (*Pan troglodytes* or *Gorilla gorilla*).

4.7×10^{-6} . The weighted values, $E(n)$, given in Table 1 are the estimate of the expected value ($2n - 6$). Small values of $d(T_i, T_j)$ are very rare.

We have at this point established: (1) a method that can guarantee to find all minimal and near minimal trees, (2) a quantitative method of comparing trees and (3) a method of associating frequencies with the comparison of trees. The next stage is to apply these methods to comparative biological data.

Application to sequence data

When this study began there were polypeptide sequences available that were common to 11 taxa. The proteins are cytochrome c, haemoglobin A, haemoglobin B, fibrinopeptide A and the last 13 amino acids of fibrinopeptide B. The original data are from the *Atlas of Protein Sequence and Structure*²¹, together with its supplements. There are cases, such as kangaroo, where two species of the same genus have been used (*Macropus congruus* and *Macropus giganteus*), but this does not affect the result. Table 2 lists the species. Our methods will work with more species and/or more sequences (for example, myoglobin, α -crystallin A, RNase), but we frequently find just one sequence is not available. There is an urgent need for more coordination in selecting sequences for analysis.

The protein sequences were converted to nucleotide sequences²² (with haemoglobin B, some nucleotide sequences were available²³). The sequence data were edited to remove invariant sites and other sites with no comparative information, as these have no effect on the structure of the phylogenetic trees (refs 10, 16 and D.P., in preparation). This left only a small number of sites, particularly for cytochrome c, as it shows little variation among these taxa. The final data are listed in Table 2, with 40 sites from haemoglobin A, 43 from haemoglobin B, 13 from fibrinopeptide A, 14 from fibrinopeptide B and 13 sites from cytochrome c.

Evolutionary tree construction

For 11 taxa there are $17!! = 34459425$ unrooted binary trees to be considered. The branch and bound algorithm was applied to each of the five protein sequences individually, as well as to the combined sequences. Table 3 lists the numbers of trees close to minimal length for each of the five data sets and for

the combined data. The 39 trees whose lengths are within 1.25% of the minimal lengths were selected for detailed comparisons and are illustrated in Fig. 1. In order that they should be presented as rooted evolutionary trees, the marsupial (kangaroo) was selected as determining the root or ancestor of the tree.

Tree comparisons

It can be seen from Fig. 1 that there are only two identical trees among these 39, $T_8 = T_{38}$. Using the comparison algorithm outlined above, we obtain the ${}_{39}C_2 = 741$ values of $d(T_i, T_j)$ ranging from 0 (1 value) to 14 (8 values) out of the maximum of 16, with a mean value of 7.57. Interpolating from the frequencies in Table 1, we would expect such a value to occur between a pair of random trees with probability 1.9×10^{-4} . There are, of course, 741 comparisons but they cannot be considered to be independent values. The average similarity for comparisons of trees from the same sequences is 4.1 and for comparisons of trees from different sequences is 8.33. The expected and observed values of m are as follows:

$m =$	0	2	4	6	8	10	12	14	16
Exp.	0	0	0	0	0	2	17	125	597
Obs.	1	53	87	163	200	145	84	8	0

Table 3 The minimal length and number of trees close to minimal length for each of the five sequences individually and combined

Sequence	Minimal length	min	(min + 1)	(min + 2)	(min + 3)	(min + 4)
Cytochrome c	17	6*	55	195	321	368
Fibrinopeptide A	29	1*	37	403	2,724	12,449
Fibrinopeptide B	36	8*	126	475	1,313	4,660
Haemoglobin A	89	1*	5*	26	143	400
Haemoglobin B	124	3*	4*	25	41	105
Combined sequences	308	2*	3*	6*	0*	8

* Those 39 trees with length $\leq 1.25\%$ of the minimal. These 39 trees were used in the comparative analysis and are shown in Fig. 1.

Table 4 The average distance between trees derived from two sequences

	CS	Cc	FA	FB	HA	HB
CS	2.0/3.9	7.0	8.0	9.8	8.0	5.3
Cc	5.8	3.1	8.3	10.2	6.0	5.7
FA	6.9	8.3	—	4.8	10.0	11.3
FB	8.2	10.2	4.8	4.5	8.8	11.8
HA	8.4	7.6	10.3	11.2	—/4.7	6.7
HB	6.5	7.6	11.1	11.5	9.0	2.7/4.4

The values in the upper triangle are from comparisons among the 21 minimal trees, the values in the lower triangle are from comparisons among the 39 trees studied. CS = combined sequences, Cc = cytochrome c; FA, FB = fibrinopeptides A and B; HA, HB = haemoglobins A and B.

There is thus a strong divergence away from random towards the trees being very similar. Note that it is logically possible for the trees to have been more dissimilar than expected.

Table 4 gives the mean values of $d(T_i, T_j)$ between trees of different sequences, the upper values being obtained only from the minimal trees, and the lower values obtained from all 39 trees. The largest mean that occurs is 11.75 between the minimal length trees of fibrinopeptide B and haemoglobin B. This value would occur between a pair of random trees with probability 1.8×10^{-2} . All other values have probabilities less than this, ranging down to 1.1×10^{-5} for fibrinopeptides A and B.

Clearly, we can reject any idea that the trees from the different sequences are independent. The different protein sequences give trees that are markedly similar, showing a relationship between them that is consistent with the theory of evolution. This supports the theory but of course does not prove it; scientific theories are falsifiable but not provable^{1,24}.

Consensus tree

Of the 34459425 unrooted binary trees for 11 taxa, we have selected the 39 near minimal trees for further study. The minimal trees with the combined sequences are one estimate for the most likely tree from these data, but it is also possible to derive a 'consensus tree' which incorporates the most common features of the 39 trees.

We can, by using the partitions, find a particular tree T such that the sum $d(T, T_1) + d(T, T_2) + \dots + d(T, T_{39})$ is minimal. In calculating the tree distances, we computed the number of times each particular partition occurred in a given tree. The nine most frequent partitions and the number of trees in which they occur are {H, A} (39); {H, A, R} (37); {S, C} (30); {S, C, P, E} (28); {S, C, P} (22); {K, D} (21); {H, A, R, O, M} (16); {E, P} (12); {O, M} (12) (see Table 2 legend for definition of abbreviations). For example, the first partition is {H, A} which is human and ape, and this occurs in all 39 trees. Each partition is defined by listing the smaller of the two subsets, and so in this case ({H, A}) the remaining nine taxa are in the other subset. The tenth most frequent partition occurs in only seven trees.

The partition {E, P} with frequency 12 is inconsistent with partition {S, C, P} with frequency 22. If we delete {E, P}, the remaining 8 partitions uniquely define the 8 links of a binary tree T spanning the 11 taxa. This tree is, in fact, the tree T_7 in Fig. 1 of length 310 on the complete data and with an average distance of 5.5 to the other trees. We call this tree the consensus tree of the set of trees and it could be regarded as being representative of the set. It is the tree which will give the minimal sum of m values when comparisons are made with all 39 trees. It is a markedly better tree on this test than any other. Other consensus tree methods are available²⁵ but were less appropriate for this application.

The consensus tree does have the advantage over the two combined sequence minimal trees (T_1, T_2) in that in the ungulates, the horse (perissodactyl) is separated from the three artiodactyls (cow, sheep and pig). References to results of mammalian phylogeny from palaeontological evidence can be found elsewhere^{26,27}. In McKenna's published scheme²⁶, lagomorphs (rabbit) would be the first branch after the marsupials. This does not occur in any of our 39 trees.

Conclusion

The general conclusions from the present work are that (1) it is possible to make falsifiable predictions from the hypothesis that species have been linked in the past by an evolutionary tree and (2) there is strong support from these five sequences for the theory of evolution. There may be exceptions where different sequences will lead to different trees as, for example, in the serial symbiosis theory²⁸. Also, in pre-cellular evolution a network with circuits may be a better model than a tree²⁹. Note also that this work has so far been confined to the question of the existence of an evolutionary tree and has not discussed the mechanism of evolution. The work can be extended by using criteria of optimality that assume particular mechanisms of evolution.

An interesting philosophical question would arise if the results of this work had falsified the prediction that the trees would be similar. Would this disprove the theory of evolution, or could it just mean that the sequences had changed so rapidly that they had lost all information about their early history, thus contradicting the hypothesis of Zuckerkandl and Pauling⁸? It could be argued that because proteins from different species can be aligned so readily, this in itself is independent evidence that the proteins retain evolutionary information. However, it is probably true that specific predictions from hypotheses, rather than the hypotheses themselves, are falsifiable. This idea is inherent in Popper's writing, but is more clearly expressed by Lakatos³⁰. To this extent, we suggest that Popper's criticisms of evolutionary theory have shown incompleteness in the application of evolutionary theory, but at the same time evolutionary theory has helped clarify some inadequacies in Popper's model of the growth of knowledge.

Received 12 October 1981; accepted 16 March 1982.

1. Popper, K. *Unended Quest: An Intellectual Autobiography* (Fontana, London, 1976).
2. Popper, K. *Dialectica* **32**, 339–355 (1978).
3. Halstead, B. *New Scientist* **87**, 215–217 (1980).
4. Ruse, M. *New Scientist* **89**, 828–830 (1981).
5. Editorial *Nature* **290**, 75–76 (1981).
6. Harary, F. *Graph Theory* (Addison-Wesley, Reading, Massachusetts, 1969).
7. Carré, B. *Graphs and Networks* (Clarendon, Oxford, 1979).
8. Zuckerkandl, E. & Pauling, L. *J. theor. Biol.* **8**, 357–366 (1965).
9. Dayhoff, M. O. & Eck, R. V. *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Springs, Maryland, 1966).
10. Fitch, W. M. *Am. Nat.* **111**, 223–257 (1977).
11. Goodman, M., Czelusniak, J., Moore, G. W. & Romero-Herrera, A. E. *Syst. Zool.* **28**, 132–163 (1979).
12. Mickevich, M. F. *Syst. Zool.* **27**, 143–158 (1978).
13. Cavalli-Sforza, L. L. & Edwards, A. W. F. *Evolution* **21**, 550–570 (1967).
14. Feisenstein, J. *Syst. Zool.* **27**, 27–33 (1978).
15. Hendy, M. D., Foulds, L. R. & Penny, D. *Math. Biosci.* **51**, 71–89 (1980).

16. Foulds, L. R. & Hendy, M. D. *J. molec. Evol.* **13**, 127–150 (1978).
17. Hendy, M. D. & Penny, D. *Math. Biosci.* **59** (in the press).
18. Smith, T. F. & Waterman, M. S. *Am. Math. Mon.* **87**, 552–553 (1980).
19. Robinson, D. F. & Foulds, L. R. *Springer Lect. Notes Math.* **748**, 119–126 (1979).
20. Waterman, M. S. & Smith, T. F. *J. theor. Biol.* **73**, 789–800 (1978).
21. Dayhoff, M. O. *Atlas of Protein Sequence and Structure 1972* (National Biomedical Research Foundation, Silver Springs, Maryland 1972).
22. Penny, D., Hendy, M. D. & Foulds, L. R. *Biochem. J.* **187**, 65–74 (1980).
23. van Ooyen, A. *et al. Science* **206**, 337–344 (1979).
24. Popper, K. R. *Objective Knowledge* (Oxford University Press, 1972).
25. Margush, T. & McMorris, F. R. *Bull. Math. Biol.* **43**, 239–244 (1981).
26. McKenna, M. C. in *Phylogeny of Primates* (eds Luckett, W. P. & Szalay, F. S.) 21–46 (Plenum, New York, 1975).
27. Szalay, F. S. in *Major Patterns in Vertebrate Evolution* (eds Hecht, M. K., Goody, P. C. & Hecht, B. M.) 315–374 (Plenum, New York, 1976).
28. Schwartz, R. M. & Dayhoff, M. O. *Science* **199**, 395–403 (1978).
29. Eigen, M. & Winkler-Oswatitsch, R. *Naturwissenschaften* **68**, 217–228 (1981).
30. Lakatos, I. in *Method and Appraisal in Physical Science* (ed. Howsen, C.) 1–40 (Cambridge University Press, 1976).

DNA sequence of a foldback transposable element in *Drosophila*

S. Steven Potter

Department of Biology, Wesleyan University, Middletown, Connecticut 06457, USA

The bizarre structures of the Drosophila foldback (FB) transposable elements are discussed here, and arguments favouring a transposase-mediated mobility are presented. The complete 4,089-base pair nucleotide sequence of transposable element FB4 reveals the full pattern of periodicity within the inverted repeats and the nature of the loop-inverted repeat junctions, as well as the coding potential of the loop sequence.

TRANSPOSABLE elements are discrete units of DNA that are capable of occupying new genomic locations; they can covalently insert into different positions in the genome. Since their discovery¹, advances in technology have allowed the molecular isolation and characterization of a wide variety of transposable elements from bacteria and eukaryotes²⁻⁹ (for reviews see refs 10, 11).

Here we present the complete nucleotide sequence (4,089 base pairs, bp) of the FB4 transposable element from *Drosophila melanogaster*. We have previously shown that the FB family of mobile sequences is distinct from the retrovirus-like transposable elements, such as *copia*^{12,13}. The FB elements represent the only eukaryotic transposable elements discovered which have large inverted terminal repeats. Moreover, these FB terminal repeats vary in size from one element to the next, and many FB elements consist entirely of contiguous inverted terminal repeats, with no detectable separating DNA.

The FB4 nucleotide sequence is of interest for the following reasons. First, the data reveal the structural details of the periodicities found within the inverted repeats. Second, comparison of the two inverted repeats of FB4 shows differences that reflect the nature of some of the genetic events associated with these elements. Third, the loop sequence of FB4, separating the inverted repeats, is quite different from that found on any of the other cloned FB elements. Sequence analysis yields the structure of the loop-inverted repeat junctions and reveals the coding potential of the loop sequence.

The restriction map of FB4 and the sequencing strategy used are shown in Fig. 1. The complete base sequence of FB4 is presented in Fig. 2.

Inverted repeat structure

Each inverted repeat of FB4 is constructed primarily of small direct repeats. Near the distal termini of the inverted repeats

there are multiple imperfect copies of the 10-bp sequence CGTTTGCCCA. These short repeats are, however, generally separated by large regions of more diverse DNA. At nucleotide 222 this 10-bp repeat is expanded to give 20 base pairs, CGTTTGCCCAACCCTTTAAAA. This 20-bp repeat is also present in multiple imperfect copies, separated by variable regions of A+T-rich DNA. Starting at 499 bp from the end we find 31-bp contiguous tandem repeats which contain within them the 20-bp repeat. These 31-bp repeats continue for the ~500 bp remaining in the inverted repeats.

Figure 3 illustrates schematically this unusual sequence construction. The thick arrow represents one of the inverted repeats of FB4, and the shorter arrows the 10-bp repeats found near the distal termini. Figure 3 is only schematic, thus there are more copies of the various repeats than those shown.

Figure 4 illustrates the periodicity found within the tandem repeats of the 31-bp sequence. Not all copies of this repeat are identical. Five types of 31-bp repeat together make up a 155-bp (5×31) repeat that is itself tandemly repeated. In Fig. 4 the sequence from the proximal portions of the inverted repeats is shown with the 31-bp repeats in register and labelled according to type. The periodicity, with every fifth repeat being of the same type, is apparent.

This unusual sequence organization, with 10-bp repeats within 20-bp repeats within 31-bp repeats within 155-bp repeats within inverted repeats, leads us to speculate on the possible function. We propose the following scheme. Transposase levels in the eukaryotic cell, even in optimum conditions, may be extremely low, and the termini of an individual transposable element represent a minute fraction of the large genome. Therefore the rate-limiting step in the transposition process may be the recognition and binding of transposase to the ends of the transposable element, where the enzyme has its effect. The tandem repeats may serve to increase the size of the target



Fig. 1 Sequencing strategy. The arrows underneath represent sequencing experiments using the existing restriction recognition sites shown. Because of the repetitious nature of the inverted repeats there are relatively large regions of FB4 having few restriction sites. We therefore introduced new restriction sites by mutagenesis *in vitro*. The purified plasmid pDmFB4 (for a complete restriction map see Potter *et al.*¹²) was cleaved with *Xba*I and then treated for varying periods of time with the exonuclease *Bal*31. The reaction was monitored by gel electrophoresis. Synthetic polynucleotide linkers carrying *Xba*I recognition sites were blunt-end-ligated to the plasmid with T4 ligase. The DNA was then digested with *Xba*I, circularized by a second round of ligation, and used to transform *Escherichia coli* K-12 HB101. Ampicillin-resistant colonies were selected and screened by restriction analyses. The arrows above the map of FB4 represent sequencing experiments using the restriction sites introduced by *in vitro* mutagenesis. Enzymes and polynucleotide linkers were from New England Biolabs; isotope was from NEN. Sequencing was by the procedure of Maxam and Gilbert²⁸. The five reactions—G, GA, CT, T and A—were used. End-labelling was performed as described previously^{13,22,28}.

1 AGCTCAAAGAAGCTGGGGTCGGAAAAATCGAATTTTGAATTTGAAAGCTGGAATCGTTTGGCCATTTTGGCCCATGTTGCCACCAATTAGTTTTT
 101 TTTGCCACGTCAGTTTTTGAGATATGGATTTTCGAAAAGTTGCAAAATGTTGCAAAATCAAAATTTTCGCTTTTTCAAATTTTTCCTTTTAAAT
 201 CGCAATAACATCGTTTGGCCACGTTTGGCCACCCCTTTAGAAATTTGAAAAATTTATACCTTTAGAAAATATAAGGCTTTTAAGTTTACCTCGGTCTAATC
 301 AGAGAGTAAATCGTTTGGCCATCTCTTAAACCAAAATTTATCAACAAAAACGTTTGGCCAAACCATTTATTAGTTTTCGTTTGGCCACCCCTTTA
 401 AAAAACTTTTACAAAAATTTTTCGATTGGCCACACTTGAATACAAACCAATTCGTTAGCCACCTCTTCAAAATAAATATTCCAATAAAAAACG
 501 TTTTCCACCATTAAAAATAAATTTTCGATTGGCCACCTCTCAAAATTCATTTTAACGTTTGGCCACCCCTTAAAAATTTGTTTTTTCGTTTGGCCCA
 601 CTCCTAAAACTAAATAATTTTCGATTGGCCACCTTTAAAACTAAATAATTTTCGTTTGGCCATCCTTTAAAAATTCATTTTAACGTTTGGCCACCCCTTTAA
 701 AAATAAATTTATTCGTTTGGCCACCCCTTAAAAATTTGTTTTTTCGTTTGGCCACTCTTAAAACTAAATAATTTTCGATTGGCCACCTTTAAAACTAAAT
 801 AATTTTCGTTTGGCCACCTCTTAAAAATTCATTTTAACGTTTGGCCACCCCTTAAAAATAAATTTTCGTTTGGCCACCCCTTAAAAAGTTTTTTTTTCG
 901 TTTGCCACCTCTTAAAACTAAATAATTTTCGATTGGCCACCCCTTAAAACTAAATAATTTTCGTTTGGCCATCCTTTAAAAATTCATTTTAAACGTTTGGCCA
 1001 CCTTTAAAAATTTGTTTGAAGATGTGGGCCCAATTCAGATATTTTAGGATCGGCGGATAGAAGCACTTACTATATGATGATGATGAACATACATAGA
 1101 CATAATATGTACAGCTGTGTTTCAGAAAAATAGCAGTGGCAAGGAACTAAGTAATACAAAGGTATTTTCCATGTCCCTTTTCGGAATCGACTTTTATT
 1201 CCTCTTATTTTGTAAATGGAATGTGTAGATAGGAAAAAAGAAAAATCCGGTCAGTTTCTTGTATCTCTTTTATTTATTCATCTTGGAGCAAAAT
 1301 CACAATTTTATAGCTGTTTATAAAGATAGCAGTGTCTGTTCTGACCAACGTAAGATCCCGAAATCATCAATTTTCTAAAAAGTGAGTTTGGTTAAGT
 1401 TAATTCGTATATTTAAAGGCACTAATAATTTAAAAAATTTTATTTTATAGTGGTATAGGACAGCACTACTCCAGGGGAAAAAGATGTTAAT
 1501 TCTTAAGCTTGAAGAAGGAAAAACATATAAGGACATCAAAAAACCCCTTAAATGTTCTGCAAAATGGTATCCCATGCCATTAATAATAAATGGAAG
 1601 CCCGAAAACCGTGTACCAAACTAAATACACAGATATAGGAGTGCAGCAGTGTCTTCTTACAGCAAGTCTATCGTTTTCATCCTTTAGGGACATAA
 1701 AGTCTGAGCTGAICTTGGGAATCAGCGACTTACTATTCTGAGAGCTACTGAATCAAAATTTTCAGTGGCAGGAGTCCAGAAAGTTCCCTACCTAG
 1801 CCCAAGGCATATTAAGCAAGGTTAAGCTTAGCTTAAACCTACCTAACTGGCCAGTCTCCAAATGGCGTAATATCTTTTGGAGTATGGGTCAAAATC
 1901 ATGCTATTGTTGGTGAAGCTGTTCTACTACAGTATATCTGACGACCTCCAAACCGAGTATCCCAAAACCCAGTGAAGACTTTCAATCAGCGTGGAC
 2001 CTAAATCATGTTATGGGCTGTTTTTTTATAATGTTATGAGTATGCTATGCTATGATTATGTTATGATTATAGACCAAAACGCATATGTAATATAC
 2101 TTAGTGATGTTCTTATGTCATATTTCTGAATAAATATACCTTAAATGGCATTCCAACAGGATAATGATCAGAAACGCAGATGTAATCGGCTAAGAA
 2201 TAGGTTCAACCAAAATAGATAGATGCAATGCGGTGCAAGCACCCTTCCCATTTAAACCCGATGAAACCTGTATGGGACATTAACAGTTTGTG
 2301 TCGAAGAAGTCCCGACGCTTAAGACTCAGATTTGGCAAGTTGTGCAAGATACATGGGCAAAATTCCTCCCAACCTTGTAGGACTTGGTGGACTTCA
 2401 TGCGCGTGGGTGAAGGCTGTCTGGCTAACAAAGGCTATCCAGCCAAGTATTAGGCCGAATTAACATATTAAAAAGAAAACTAAGTCTGTTCTAGG
 2501 TCAAGTTAAATTTGTTACTATTTTCTATAGCACTGCTATTTTATGAACACCAAGTTTCTGCCTATTATTGTTTAACTATATTTTCGAAACTAT
 2601 TGAAGAAATAAAGTGAAACATTGTTTAAATTTGTTGAAATGAATACCTAATGATATTATTAATAAAAAATTCCTTAAAACTGTAATCATAGGAAT
 2701 TTTTATCTTAAACTCTGAAGTCCAAAGCACTGCTATTATTTCTGAACACAGCTGTACATAACGAAATAATTTTAAAGGTTGGGCAACGTTAAAA
 2801 ATGAATTTAAAGGATGGCAATCGAAATTTATTTAGTTTAAAGATGGGCAACGAAAAACAAATTTTAAAGGTTGGGCAACGAAATACTTTATTTT
 2901 TAAAGGTTGGGCAACGTTAAAAATGAATTTTAAAGGATGGGCAACGAAATTTATTTAGTTTAAAGGTTGGGCAATCGAAATTTATTTAGTTTAAAGAT
 3001 GGGCAACGAAAAACAAATTTTAAAGGTTGGGCAACGAAATTTATTTTAAAGGTTGGGCAACGTTAAAAATGAATTTTAAAGGATGGGCAAA
 3101 CGAAATTTATTTAGTTTAAAGGTTGGGCAATCGAAATTTATTTAGTTTAAAGATGGGCAACGAAAAACAAATTTTAAAGGTTGGGCAACGAAATAC
 3201 TTTATTTTAAAGGTTGGGCAACGTTAAAAATGAATTTTAAAGGATGGGCAACGAAATTTATTTAGTTTAAAGGTTGGGCAATCGAAATTTATTTAGTT
 3301 TTAAGATGGGCAACGAAAAAATCTTTTAAAGGTTGGGCAACGAAATTTATTTTAAAGGTTGGGCAACGTTAAAAATGAATTTTAAAGGAT
 3401 TGGGCAACGAAATTTATTTAGTTTAAAGGTTGGGCAATCGAAATTTATTTAGTTTAAAGATGGGCAACGAAAAACAAATTTTAAAGGTTGGGCAAA
 3501 CGAAATAATTTATTTTAAAGGTTGGGCAACGTTAAAAATGAATTTTAAAGGATGGGCAACGAAATTTATTTAGTTTAAATGTTGGGCAACGTTT
 3601 TTTTGGAAATTTTATTTTAAAGGTTGGGCAACGAAATTTGGGTGATTTTAAATGTGGGCAATCGAAAAAATTTTGTAAAGGTTTAAAGGG
 3701 TGGGCAACGATAAACTAATAAATGTTGGGCGTTTTTGTGATAATTTTGGTTTAAAGATGGGCAACGATTACTCTCTGATTAGACCGA
 3801 GGTAACTTAAAGCCTTATATTTTCTAAGATATAATTTTCTAAAAATCTAAGGTTGGGCAACGTTGGGCAACGATGTTATTCGATTAAAAAA
 3901 AAAAATTTTGAAGAAAGCAATTTTGTATTTTCGAAATTTTTCGAACTTTTTCGAAATTCATATCTCAGAACTGGACGTGGGCAAAAAAATAAT
 4001 GGTGGGCAACATGGGCAAAAAATGGGCATACGATTCAGCTTTCAAAATTTCAAAATTCGATTTTTCGACCCAGCTTCTTGTAGCT 4089

Fig. 2 Nucleotide sequence of the FB4 transposable element. The complete base sequence of FB4 is shown, starting with the first nucleotide of the left inverted repeat. The various imperfect copies of the 10- and 20-bp repeats within the left inverted repeat are underlined, as are the punctuation signals surrounding the open reading frame of the loop and the 33-bp inverted terminal repeat of the loop. CAT, the CAT box; G-H, the Goldberg-Hogness box; poly(A), the polyadenylation signal.

recognized by the transposase. In a two-stage binding process, the transposase would first attach loosely but specifically at the tandem repeats and then move randomly (although a more directed movement is possible) until encountering the FB termini, where tighter binding and catalytic activity would occur.

This 'bind and slide' model, although conjectural, does provide testable predictions. For example, other factors being equal, an FB element having larger inverted repeats (and hence a larger transposase target size) will undergo more frequent transposition. In addition, the model suggests an affinity column transposase purification procedure, using the tandem repeats for selective binding.

There is a precedent for the proposed two-stage recognition. For example, the *lac* repressor apparently binds loosely to any DNA sequence, moves randomly along the DNA until it locates the specific recognition site, and then binds tightly¹⁴.

Comparison of inverted repeats

Although the two inverted repeats of FB4 are very similar, there are, nevertheless, many single base differences and more

dramatic differences involving entire copies of the 31-bp sequence. There are two 31-bp repeats missing in the left inverted repeat (Fig. 4) which are present at corresponding positions in the right inverted repeat. Differences of this sort could be generated by unequal cross-over events. Given the structure of the FB elements, both interchromosomal and intrachromosomal unequal cross-overs are possible. An intrachromosomal event, involving the two inverted repeats of a single FB4, would delete sequence from one inverted repeat and transfer it to the other. Because there are no extra copies of the 31-bp sequence at the corresponding positions on the right inverted repeat, we conclude that interchromosomal events generated these particular deletions.

Intrachromosomal recombinations between the inverted repeats of a single FB element would also result in an inversion of the loop sequence and part of the inverted repeats. Such inversions in analogous structures in *Salmonella*¹⁵ and *Mu* phage¹⁶ have been related to the regulation of gene expression. It is tempting to consider that such gene control mechanisms may also exist in eukaryotes.

Because the FB elements afford multiple possibilities for out-of-register pairing, they may represent 'hotspots' for normal homology-dependent recombination events. In this way they would contribute to genetic diversity by simply accelerating the regular recombination process at selected points along the chromosomes.

Further comparison of the two inverted repeats of FB4 reveals a missing 31-bp sequence in the right inverted repeat (Fig. 4). Note also that the right inverted repeat structure is significantly larger, with approximately two extra copies of the 155-bp repeat. Our restriction mapping data for several FB members suggest that such a difference in size of the 'inverted repeats' of a single FB element is common. Indeed the 'loop' sequences, as observed by electron microscopy, are apparently often the result of this size difference. That is, the 'extra' tandem repeats from one of the inverted repeats can form the loop.

The FB4 element, however, represents a striking exception in that the bulk of the relatively large loop appears by restriction mapping to be unrelated in sequence to the inverted repeats.

The FB4 loop

The 31-bp repeats terminate abruptly at the proximal junctions with the loop. This is in marked contrast to their gradual development from the distal ends, where 10-bp repeats become 20-bp repeats, which in turn become 31-bp repeats.

Directly abutting the proximal end of the right inverted terminal repeat is a 33-bp sequence (unrelated to the 31-bp repeat) that is also found in a nearly perfect (32/33 bp) inverted repeat near the other inverted repeat-loop junction. That is, a 33-bp sequence at the right inverted repeat-loop junction is also found, in an inverted orientation, 95 bp from the left inverted repeat-loop junction. The significance of this observation, if any, is unknown. The presence of such terminal repeats suggests that the loop itself may be a transposable element, and that FB4 is a composite structure, consisting of a mobile element (loop) within a mobile element (FB).

One of the signal characteristics of a transposable element is the presence of short direct flanking sequences that result from the duplication of target site DNA during the insertion process. The *PvuII* recognition sites near the loop-inverted repeat boundaries are within 8-bp palindromes (ACAGCTGT) which can be regarded as either part of the full 33-bp inverted repeats of the loop, or as direct repeats flanking a shorter version of these inverted repeats (now only 20 bp long). Further support for the loop being a distinct transposable element is provided by preliminary results which indicate that these loop sequences are present at numerous genomic positions, where they are not always associated with FB elements (S.S.P., unpublished observations).

Alternatively, the loop may represent an integral and natural component of the FB4 element. This explanation avoids some of the difficult questions posed by the 'transposable element within a transposable element' model. For example, why is there a 95-bp space between the left inverted repeat-loop junction and the 33-bp 'terminal' repeat of the loop? And why did the loop transposable element insert into the centre of the FB element? (If it inserted within one of the inverted repeats, then a polarity switch would be seen in the tandem repeats of

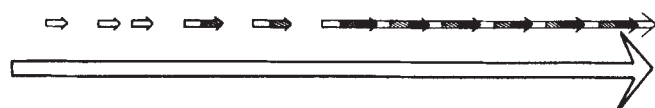


Fig. 3 Construction of the inverted repeats. The short arrows represent the 10-bp repeats, which expand to form 20-bp repeats which expand again to form 31-bp tandem repeats within the inverted repeat. There are many more copies of these various repeats than shown in this diagram.

Left	Right
1. CGTTTCCCACCATTTAAAAATAAATTT	1. CGTTTGCCACCATTTAAAACTAAATAATTT
2. CGATTGCCATCCTTTAAAATTCATTTT_AA	2. CGTTTGCCATCCTTTAAAAATTCATTTTAA
3. Absent	3. CGTTTGCCACCCCTTTAAAAATAAATTTATTT
4. CGTTTGCCACCCCTTTAAAAATTTGTTTTTTT	4. CGTTTGCCACCCCTTTAAAAATTTGTTTTTTT
5. CGTTTGCCACTCTTTAAAACTAAATAATTT	5. CGTTTGCCACTCTTTAAAACTAAATAATTT
1. CGATTGCCACCTTTTAAAACTAAATAATTT	1. CGATTGCCACCTTTTAAAACTAAATAATTT
2. CGTTTGCCATCCTTTAAAAATTCATTTTAA	2. CGTTTGCCATCCTTTAAAAATTCATTTTAA
3. CGTTTGCCACCCCTTTAAAAATAAATTTATTT	3. CGTTTGCCACCCCTTTAAAAATAAATTTATTT
4. CGTTTGCCACCCCTTTAAAAATTTGTTTTTTT	4. CGTTTGCCACCCCTTTAAAAATTTGTTTTTTT
5. CGTTTGCCACTCTTTAAAACTAAATAATTT	5. CGTTTGCCACTCTTTAAAACTAAATAATTT
1. CGATTGCCACCTTTTAAAACTAAATAATTT	1. CGATTGCCACCTTTTAAAACTAAATAATTT
2. CGTTTGCCATCCTTTAAAAATTCATTTT_AA	2. CGTTTGCCATCCTTTAAAAATTCATTTTAA
3. CGTTTGCCACCCCTTTAAAAATAAATTTATTT	3. CGTTTGCCACCCCTTTAAAAATAAATTTATTT
4. CGTTTGCCACCCCTTTAAAAATTTGTTTTTTT	4. CGTTTGCCACCCCTTTAAAAATTTGTTTTTTT
5. CGTTTGCCACTCTTTAAAACTAAATAATTT	5. CGTTTGCCACTCTTTAAAACTAAATAATTT
1. CGATTGCCACCTTTTAAAACTAAATAATTT	1. CGATTGCCACCTTTTAAAACTAAATAATTT
2. CGTTTGCCATCCTTTAAAAATTCATTTTAA	2. CGTTTGCCATCCTTTAAAAATTCATTTTAA
3. Absent	3. CGTTTGCCACCCCTTTAAAAATAAATTTATTT
4. CGTTTGCCACCCCTTTAAAAATTTGTTTTTTT	4. CGTTTGCCACCCCTTTAAAAATTTGTTTTTTT
	5. CGTTTGCCACTCTTTAAAACTAAATAATTT
	1. CGATTGCCACCTTTTAAAACTAAATAATTT
	2. CGTTTGCCATCCTTTAAAAATTCATTTTAA
	3. CGTTTGCCACCCCTTTAAAAATAAATTTATTT
	4. CGTTTGCCACCCCTTTAAAAATTTGTTTTTTT
	5. CGTTTGCCACTCTTTAAAACTAAATAATTT
	1. Absent
	2. CGATTGCCATCCTTTAAAAATTCATTTTAA
	3. CGTTTGCCACCCCTTTAAAAATAAATTTATTT
	CGT

Fig. 4 Periodicity of the 31-bp repeats. The sequences of the proximal portions of the inverted repeats are shown. Each type of 31-bp repeat is labelled by a number. The periodicity of types is apparent, although in three cases an expected copy of the 31-bp repeat is absent. To facilitate comparison, the reverse complement of the right inverted repeat is shown: the left inverted repeat sequence illustrated starts with nucleotide 499 and ends at position 1,018, while the right inverted repeat sequence shown starts with base 3,595 and ends at 2,761. Underlined bases deviate from the standard sequence of that repeat type. Underlined blank spaces represent a missing nucleotide that is normally present.

the other side of the FB element.) The integral component model is also attractive because it allows the FB4 element greater sequence complexity. The inverted repeats of the FB elements, although sometimes over 1 kb long, consist of repetitious sequences with frequent translational stop signals in all three possible reading frames, and it is therefore unlikely that they encode proteins. Where, then, is the transposase responsible for movement of these FB elements encoded? Loop sequences are obvious candidates, especially as bacterial transposons can carry transposase genes within their loops¹⁷.

The loop sequence of FB4 contains a single large open reading frame, with the methionine codon beginning at nucleotide 1,494 and the next stop codon, TGA, beginning at nucleotide 1,938, allowing a maximum of 148 amino acids to be encoded. The consensus sequence GG^CCAATCT, termed the

CAT box, is often found upstream from eukaryotic genes. At the appropriate distance upstream from the coding region there is the sequence GATCAATAT which agrees with the CAT box for seven out of nine bases. There is also a good Goldberg-Hogness box and after the open reading frame, a perfect copy of the sequence AATAAA which is generally found 14–30 nucleotides upstream from the poly(A) tract of most mRNAs. It seems reasonable, therefore, to conclude that this open

reading frame sequence is a real gene, although the function of the gene product remains unknown.

It is also possible that other regions of FB4 encode small polypeptides, or that RNA splicing allows them to encode more substantial proteins.

Mechanism of transposition

Two distinct modes of duplication-transposition are now clear in outline, if not in detail. The bacterial transposons apparently move strictly via DNA intermediates, as proposed, for example, by the Shapiro model¹⁸. The finding of 'cointegrate' transposition intermediates¹⁰, in which the two transposons are still linked, before the resolution event, strongly supports this general type of model.

The integrated proviral DNA copies of retrovirus RNA genomes undergo a more complicated duplication-transposition process. The proviral genome is transcribed into RNA which is packaged into a viral particle and eventually enters another cell, where it is reverse-transcribed¹⁹, and a circular DNA intermediate having one or two copies of the long terminal repeats, is formed. This DNA circle can then integrate into the new host genome²⁰. To generate retroviral intragenomic duplication-transposition events one need only eliminate the virion stage. The proviral transcript would be immediately reverse-transcribed, circularized and reintegrated at another chromosomal location.

The *copia*-like transposable elements in *Drosophila* are indistinguishable from integrated proviral retrovirus genomes. Structurally they both carry long direct terminal repeats^{4,21,22} and the transcription patterns are very similar²³. In addition, Flavell and Ish-Horowitz have shown²⁴ that *copia* elements give rise to circular DNA intermediates having one or two copies of the long direct terminal repeat, just as the retroviruses do. It therefore seems probable that the retroviruses and retrovirus-like transposable elements such as *copia* share a common duplication-transposition mechanism involving reverse transcription.

It has also been proposed that the human snRNA pseudogenes, with their dispersed and truncated organization, might result from a reverse transcription duplication-transposition process²⁵. However, for these sequences, unlike the retroviruses which are designed to be reverse-transcribed, the process generally produces fragmentary versions of the original gene.

Jagadeeswaran *et al.*²⁶ have suggested that the human *alu* sequences could also duplicate and transpose by reverse transcription. Because these genes can be transcribed by RNA polymerase III, which recognizes internal promoters, the process is potentially 'explosive'. (The new *alu* genes can also be transcribed, giving rise to still more *alu* genes.) Furthermore, the *alu* transcripts would be capable of self priming for reverse transcriptase, giving rise to more faithful copies of the *alu* sequence.

Structurally the FB sequences are unique. They carry very large inverted terminal repeats, unlike anything found in a retrovirus, and do not seem to be designed for reverse transcription. Moreover, they have highly conserved termini, which argues against transposition by a less precise reverse transcription process, such as suggested for the *alu* and snRNA genes. Structurally the FB elements most closely resemble those bacterial transposons which have large inverted terminal repeats. As the bacterial elements move via transposase one can argue, by analogy, that the FB elements do also. Furthermore, as discussed previously, such a transposition mechanism can be invoked to help explain the unusual structure of the FB inverted repeats. The tandem repeats might serve to facilitate the binding of the low levels of transposase within the nucleoplasm. The retroviruses and retrovirus-like transposable elements, in contrast, rely on the much more abundant enzyme RNA polymerase to accomplish the onset of transposition. They therefore do not need to duplicate binding sequences.

Conclusion

The complete base sequence of FB4 reveals the full pattern of periodicity found within the inverted repeats. Near the distal termini there are occasional copies of a 10-bp sequence. Further along, this 10-bp sequence expands to 20 bp and this 20-bp repeat in turn expands to a 31-bp sequence which is tandemly repeated. There are five types of 31-bp repeat that together comprise a 155-bp repeat. This hierarchical repetition then terminates abruptly at the loop-inverted repeat junctions. Many differences exist between the left and right inverted repeats. The data suggest that these sequences are hotspots for unequal crossing-over.

The precise nature of the FB4 loop sequence remains uncertain. It is clear, however, that in this case most of the loop shows no significant homology with the inverted repeats. The presence of a gene within the loop is suggested by an open reading frame flanked by the correct punctuation signals.

It is intriguing that most FB elements cannot encode transposase. Such a dependence on transposase encoded elsewhere would seem hazardous to the transposable elements. Inactivation of the transposase gene(s) by deletion or mutation would render all the FB sequences immobile, at which point they might simply rest, accumulate mutations, and perhaps eventually be deleted themselves.

Finally, note that transposable elements interacting with transposase can generate a variety of chromosomal rearrangements, including inversions and deletions in contiguous sequences¹⁸. We have previously shown, by Southern blot experiments, the existence of several different DNA configurations within the vicinity of FB4 (ref. 12). The intriguing results of Goldberg *et al.*²⁷ also strongly suggest that two FB elements can cooperate to mobilize a large block of separating DNA. This further accentuates the many similarities found between the FB sequences and the bacterial transposable elements.

Received 5 February; accepted 26 March 1982.

- McClintock, B. *Cold Spring Harb. Symp. quant. Biol.* **21**, 197-216 (1956).
- Gafner, J. & Phillipsen, P. *Nature* **286**, 414-418 (1980).
- Pardue, M. L. & Dawid, I. B. *Chromosoma* **83**, 29-43 (1981).
- Potter, S. S., Brorein, W. J. Jr, Dunsmuir, P. & Rubin, G. M. *Cell* **17**, 415-427 (1979).
- Roeder, G. S. & Fink, G. R. *Cell* **21**, 239-249 (1980).
- Tchurikov, N. A., Zelentsova, E. S. & Georgiev, G. P. *Nucleic Acids Res.* **8**, 1243-1258 (1980).
- Cameron, J. R., Loh, E. Y. & Davis, R. W. *Cell* **16**, 739-751 (1979).
- Strathern, J. N., Spatola, E., McGill, C. & Hicks, J. B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2839-2843 (1980).
- Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280**, 288-294 (1979).
- Calos, M. P. & Miller, J. M. *Cell* **20**, 579-595 (1980).
- Spradling, A. C. & Rubin, G. M. *R. A. Rev. Genet.* **15** (in the press).
- Potter, S., Truett, M., Phillips, M. & Maher, A. *Cell* **20**, 639-647 (1980).
- Truett, M. A., Jones, R. A. & Potter, S. S. *Cell* **24**, 753-763 (1981).

- Berg, O. G., Winter, R. & von Hippel, P. H. *Biochemistry* **20**, 6929-6948 (1981).
- Zieg, J. & Simons, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4196-4200 (1980).
- Howe, M. H. *Nature* **271**, 608-610 (1978).
- Heffron, F., McCarthy, B. J., Ohtsubo, H. & Ohtsubo, E. *Cell* **18**, 1153-1163 (1979).
- Shapiro, J. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1933-1937 (1979).
- Gilboa, E., Mitra, S. W., Goff, S. & Baltimore, D. *Cell* **18**, 1383-1396 (1979).
- Shoemaker, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3932-3936 (1980).
- Finnegan, D. J., Rubin, G. M., Young, M. W. & Hogness, D. S. *Cold Spring Harb. Symp. quant. Biol.* **42**, 1053-1063 (1978).
- Levis, R., Dunsmuir, P. & Rubin, G. M. *Cell* **21**, 581-588 (1980).
- Flavell, A. J., Levis, R., Simon, M. A. & Rubin, G. M. *Nucleic Acids Res.* **9**, 6279-6291 (1982).
- Flavell, A. J. & Ish-Horowitz, D. *Nature* **292**, 591-595 (1981).
- Van Arsdell, S. W. *et al. Cell* **26**, 11-18 (1981).
- Jagadeeswaran, P., Forget, B. G. & Weissmann, S. M. *Cell* **26**, 141-142 (1981).
- Goldberg, M. L., Paro, R. & Gehring, W. J. *EMBO J.* (in the press).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).

Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus

N. Kitamura*, A. Kitamura, K. Toyoshima[‡], Y. Hirayama & M. Yoshida[†]

Cancer Institute, Toshima-ku, Tokyo 170, Japan and [‡]Institute for Medical Sciences, University of Tokyo, Tokyo 108, Japan

From the complete nucleotide sequence of the genome of the avian sarcoma virus Y73, we have predicted amino acid sequence of p90^{gag-yes}, the product of the transforming gene. Contrary to previous evidence from molecular hybridization studies, p90^{gag-yes} was found to have much homology with the transforming gene product p60^{src} of Rous sarcoma virus, suggesting that the cellular counterparts of the two (c-yes and c-src) originated from a common prototype sequence.

A GROUP of retroviruses, the acute leukaemia/sarcoma viruses, contains the transforming gene (v-*onc*), which is responsible for initiation and maintenance of cellular transformation¹. More than 13 genes have been identified as v-*onc* and shown to be acquired from a normal cellular sequence as an integral part of their genome. Three distinct classes of transforming gene capable of inducing fibrosarcomas in chicken were defined among avian sarcoma viruses (ASV), that is, *src* in Rous sarcoma virus (RSV), *yes*^{2,3} in Y73 and Esh sarcoma virus, and *fps*⁴⁻⁶ in Fujinami and PRCII strains of ASV. These three transforming genes have no homology in nucleic acid hybridization²⁻⁵, and are derived from the respective sequences in normal cells by recombination with viral sequences^{2,7,8}. Furthermore, the cellular counterparts homologous with these transforming genes have been conserved during evolution⁹ (M.Y., unpublished data) and expressed in normal cells¹⁰, suggesting some that these genes perform some crucial role in normal cells.

In the light of these observations it is surprising that the transforming gene products of all these classes have protein kinase activity phosphorylating tyrosine on target proteins¹¹⁻¹⁴, and that this enzyme activity seems to be essential for cellular transformation^{15,16}. Therefore, we have investigated the structure of these genes and the gene products in an attempt to understand not only the structural basis of the functions, but also the mechanism of biogenesis of the transforming genes.

The nucleotide sequence of *src* in Schmidt-Ruppin strain RSV (SR-RSV) has been established, and the function of the gene product (p60^{src}) and biogenesis of the *src* have been discussed on the basis of the nucleotide and amino acid sequences¹⁷. We have studied a recent isolate of ASV, Y73¹⁸, which is defective in its replication and associated with leukosis virus (YAV) of subgroup A. The genome of Y73 is 26S RNA and contains *yes* as a transforming gene in its middle portion¹⁹, which encodes polypeptide with a molecular weight of 90,000 (p90^{gag-yes}) consisting of a *gag*-gene derived part and Y73-specific portion. *In vitro* translation of 26S RNA⁴ showed directly that p90^{gag-yes} is coded by the Y73-specific RNA sequence. The association of protein kinase activity with p90^{gag-yes} suggests that p90^{gag-yes} is a transforming gene product.

To investigate the structure of p90^{gag-yes}, we determined the nucleotide sequence of the whole genome DNA. The sequence contains an uninterrupted reading frame that can code for p90^{gag-yes} and the predicted amino acid sequence of p90^{gag-yes} is strikingly similar to that of p60^{src}.

Sequencing strategy

For determination of the nucleotide sequence, we used the molecular cloning technique to amplify Y73 genome DNA. Closed circular DNA specific to Y73 was isolated from Y73-infected chick embryo cells by extraction of DNA by Hirt's procedure²⁰ followed by acid phenol extraction²¹ and agarose gel electrophoresis. The purified DNA was linearized with restriction endonuclease *Sst*I and inserted into λ phage Charon 16A²² at site of *Sst*I. Several clones containing the Y73-specific sequence were isolated and one of them, λ Y73-11A, was selected for sequencing, because it was shown by transfection assay to contain the complete sequence of Y73 RNA. Cotransfection of this cloned DNA together with cloned helper proviral DNA induced cellular transformation of chick cells and viral replication, the progeny virus was indistinguishable from the original Y73 virus. Details of the cloning the transfection will be reported elsewhere (manuscript in preparation).

Figure 1 shows a restriction map of clone λ Y73-11A insert containing two long terminal repeats (LTRs). The DNA fragments generated by digestion with *Hpa*II, *Hinf*I, and *Sau*3A were used for sequencing. The fragments were phosphorylated at the 5' end with polynucleotide kinase and [γ -³²P]ATP, separated by either strand separation or further digestion with various restriction enzymes and sequenced by the method of Maxam and Gilbert²³. Most of the sequences were confirmed by sequencing the complementary strand, or the fragments generated by digestions with different enzymes.

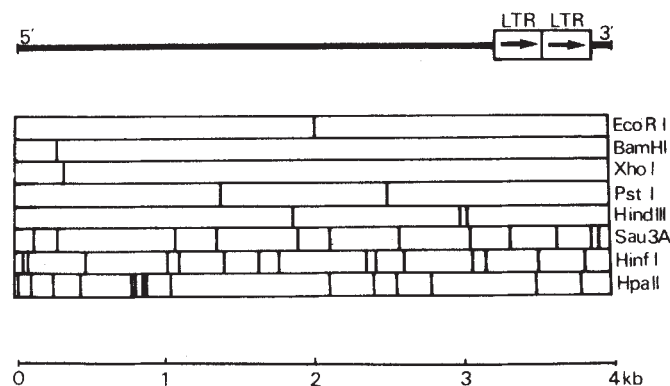


Fig. 1 Restriction map of Y73 genome cloned in λ Y73-11A. The circular DNA was linearized and inserted into λ Charon 16A vector at *Sst*I site and the cloned DNA contained two direct repeats of LTR.

* Present address: Institute for Immunology, Kyoto University, Kyoto, Japan.

[†] To whom reprint requests should be addressed.

[illegible]

Fig. 2 Total nucleotide sequence of Y73 viral genome and the predicted amino acid sequence of p90^{gag-yes}. The complete sequence of the cloned DNA in λ Y73-11A was shown in the arrange of RNA genome from 5' to 3' direction. U5 and U3 indicate the unique sequence to 5' and 3' terminal regions. Thick solid arrows with 'rec' shows the predicted recombination sites between c-yes and YAV sequences, and 'spc' represents sequence which is not contained in RSV. The horizontal thick arrows indicate the initiation and termination sites for p90^{gag-yes}. The small arrow with p19 indicates a sequence homologous to the carboxy-terminus of p19 of Pr-RSV. The closed bar between nucleotides 2,477 and 2,479 shows the putative phosphoacceptor tyrosine contained in the same nonapeptide described by Neil *et al.*²⁶ and the small open circles between 3,174 and 3,176 show the position expected as termination codon for gp37, if present.

Nucleotide sequence

Previously, fingerprint analysis of Y73 26S RNA and YAV RNA¹⁹ demonstrated that 26S RNA contained a specific sequence that represents the transforming gene of Y73. This specific sequence was mapped in the middle of the genome with sequences common to the YAV RNA at both ends. The transforming gene product p90^{gag-yes} encoded by the Y73-specific sequence was coprecipitated with antibody against p19, a gag-gene product, suggesting that p90^{gag-yes} contained at least p19 in the amino-terminal portion¹².

The entire nucleotide sequence of the Y73 genome, shown in Fig. 2, consists of 3,718 nucleotides, confirming the previous data described above. The genome was significantly shorter than 4.8 kilobases, the value estimated previously by agarose gel electrophoresis of the genomic RNA.

The noncoding sequence at the 5'-terminus, including U5 and a primer tRNA binding site, was found to be homologous with that of the SR-RSV genome^{31,32} (93% homology). The open reading frame, starting at position 380, contained the entire sequence coding for p19, which is located at the amino-

tor tyrosine in p60^{src} and p90^{gag-yes}. Our nucleotide sequence data clearly show homology over a wider range covering most of p60^{src}, thus supporting the proposal of Neil *et al.*²⁶. Some similarity between p60^{src} and the *mos* gene product of Mo-MSV was reported²⁷ and similarity was also found between two transforming gene products coded by Ha- and Ki-MSV²⁸.

Protein kinase activity phosphorylating tyrosine residues is a common feature of the transforming gene products of all avian sarcoma viruses so far tested. Therefore, it would be very interesting to know the amino acid sequences of the gene products of other avian sarcoma viruses such as Fujinami virus and PRCII.

As previously reported^{2,8}, there are DNA sequences homologous to those of *yes* and *src*, respectively in normal chick cells (*c-yes* and *c-src*), and these cellular sequences have been well conserved during evolution, being found in fish and in humans, suggesting some important function for these genes⁹. Therefore, the similarity of p90^{gag-yes} and p60^{src} at a protein level implies similarity of the gene products coded by *c-yes* and *c-src*, and suggests that *c-yes* and *c-src* diverged from a common prototype sequence at a very early stage of vertebrate evolution.

C-src was shown to be expressed in protein p60^{c-src} in normal cells^{10,11} and *c-yes* was shown to be transcribed into mRNA (M.Y., unpublished data). We have no data that could explain why different genes coding for proteins with similar structure and function are conserved and expressed in normal cells, although some speculations are possible. For example, expression of these cellular genes might depend on different types of differentiation or differences in the cellular distributions, and thus functions, of the proteins.

The structural similarity between p90^{gag-yes} and p60^{src} does not necessarily imply that the mechanisms of cellular transformation by these viruses are the same. Similarity in the carboxy-terminal domain could explain the similar protein kinase activity²⁹, but, differences in the amino-terminal domain suggest a different distribution of the proteins within cells, and thus that they phosphorylate different target proteins, since the amino-terminal domain of p60^{src} is reported to be involved in anchorage of the protein to the plasma membrane³⁰.

We thank Dr Haruo Sugano for his useful discussions. This work was partly supported by Grant-in-Aid for Cancer Research from the Ministry of Education, Science and Culture of Japan.

Received 11 January; accepted 24 March, 1982.

- Bishop, J. M. A. *Rev. Biochem.* **47**, 35–88 (1978).
- Yoshida, M., Kawai, S. & Toyoshima, K. *Nature* **287**, 653–654 (1980).
- Neil, J. C., Breitman, M. L. & Vogt, P. K. *Virology* **108**, 98–110 (1981).
- Hanafusa, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3009–3013 (1980).
- Lee, W.-H. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 2018–2022 (1980).
- Gysdael, J., Neil, J. C. & Vogt, P. K. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2611–2615 (1981).
- Shibuya, M., Hanafusa, T., Hanafusa, H. & Stephenson, J. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6536–6540 (1980).
- Stehelin, D., Varmus, H. E., Bishop, J. M. & Vogt, P. K. *Nature* **260**, 170–173 (1976).
- Spector, D., Varmus, H. E. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4102–4106 (1978).
- Spector, D., Baker, B., Varmus, H. E. & Bishop, J. M. *Cell* **13**, 381–386 (1978).
- Collet, M. S. & Erikson, R. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2021–2024 (1978).
- Kawai, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 6199–6203 (1980).
- Feldman, R. A., Hanafusa, G. & Hanafusa, H. *Cell* **22**, 757–765 (1980).
- Neil, J. C., Gysdael, J. & Vogt, P. K. *Virology* **109**, 223–228 (1981).
- Levinson, A. D., Opperman, H., Levinthow, L., Varmus, H. E. & Bishop, J. M. *Cell* **15**, 561–572 (1978).
- Powson, T. *et al. Cell* **22**, 767–775 (1980).
- Czernilofsky, A. P. *et al. Nature* **287**, 198–203 (1980).
- Itoharu, S., Hirata, K., Inoue, M., Hatsuoka, M. & Sato, A. *Gann* **69**, 825–830 (1978).
- Yoshida, M., Kawai, S. & Toyoshima, K. *J. Virol.* **38**, 430–437 (1981).
- Hirt, N. *J. molec. Biol.* **26**, 365–369 (1967).
- Zaslloff, M., Ginder, G. P. & Felsenfeld, G. *Nucleic Acids Res.* **5**, 1139–1152 (1978).
- Williams, B. G. & Blattner, F. R. *J. Virol.* **29**, 555–575 (1979).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
- Collet, M. S., Erikson, E., Purchio, A. F., Brugge, J. S. & Erikson, R. L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3159–3163 (1979).
- Hunter, T. & Sefton, B. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1311–1315 (1980).
- Neil, J. C., Gysdael, J., Vogt, P. K. & Smart, J. E. *Nature* **291**, 675–677 (1981).
- Beveren, C. V. *et al. Nature* **289**, 258–262 (1981).
- Ellis, R. W. *et al. Nature* **292**, 506–511 (1981).
- Levinson, A. D., Courtneidge, S. A. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1624–1628 (1981).
- Courtneidge, S. A., Levinson, S. A. & Bishop, J. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3783–3787 (1980).
- Czernilofsky, A. P. *et al. Nucleic Acids Res.* **8**, 2967–2984 (1980).
- Swanstrom, R., Delorbe, W. J., Bishop, J. M. & Varmus, H. E. *Proc. natn. Acad. Sci. U.S.A.* **78**, 124–128 (1981).

LETTERS TO NATURE

Magnetic fields, convection and solar luminosity variability

W. C. Livingston

Kitt Peak National Observatory, Tucson, Arizona 85726, USA

Model calculations universally indicate that heat transport from the solar interior is almost entirely by convection in the outer envelope. Spiegel and Weiss¹ have recently discussed how magnetic fields can intrude into this envelope to affect convection and thus introduce transients into the Sun's luminosity. I demonstrate here from Fourier transform spectrometer observations of Fraunhofer line asymmetry that granular convection is retarded in the presence of surface magnetism. I then present full disk observations which suggest a lessening of convection over the past 5 yr. As this time interval coincides with the rise from a minimum to a maximum of solar activity, with the presumed injection of new magnetic flux from the interior, the resolved disk and full disk observations are consistent. Finally I note the continued temporal decrease in the spectroscopic temperature of the low photosphere and consider how a change of convection and the lowering temperature may be related.

The convective process is plainly manifest at the Sun's surface by the granulation pattern. Rising bright granules, transporting excess energy which is then radiated into space, are surrounded by the cooler intergranule gas which is falling. This causes a

velocity–brightness correlation which in turn gives rise to an asymmetry in Fraunhofer lines observed near disk centre^{2,3}.

Spectral line asymmetries are traditionally represented by plotting the shifts of the bisectors of the line profile (as in Fig. 1). These bisectors have a characteristic 'C' shape, the middle of which is blueshifted in wavelength relative to the laboratory value. The upper end of the C is less blueshifted, as increased density low in the photosphere, coupled with the condition of flow continuity, results in lower convective velocities. The lower end is also less blueshifted, because of the cessation of upward velocity at the top of the granulation layer as represented by the line core. Detailed calculations by Nordlund⁴ provide a quantitative basis for the above hypothesis.

A Fourier transform spectrometer (FTS) is preferred for bisector observations; it has a symmetric response profile, is capable of giving absolute wavelengths, records many lines simultaneously and is easily fed with unfocused (integrated) sunlight⁵. In a day-long sequence of full disk 1-m FTS measurements, mid C values typically display a wavelength variance of ~0.1 mÅ, corresponding to a velocity noise of 5 m s⁻¹. For historical reasons, some of the data presented here were obtained with a 13.5-m grating spectrometer. The greatest problem in this spectrograph is thermal drift which introduces a time-dependent asymmetric instrumental profile. Velocity noise is also about 10 times worse than with the FTS, but by averaging many records the noise in the data becomes tolerable.

Figure 1 shows line bisectors for a region of high magnetic flux compared with a non-magnetic area nearby. The spatial resolution is purposely low, about 1 × 2 arc min, to average over local conditions. This record is typical of some 50 regions

studied. In all cases near the disk centre the magnetic region line bisector has a smaller curvature and is less displaced to the blue than are bisectors for quiet regions. Although Fe 5,250.6 Å is a magnetically sensitive line (Landé $g = 1.5$), the $g = 0$ line Fe 5,576.1 Å gives the same result, demonstrating that we are not troubled by an instrumental artefact arising from telescope polarization, for instance, or other conceivable effects of Zeeman splitting. Dravins *et al.*³ also find that line asymmetry is independent of g .

Figure 2a compares the full disk Fe 5,250.6 mean bisector for 1976–77 with the bisector for 1980–81. These are observations made with the 13.5-m spectrometer, the records having the worst thermal drift being eliminated, and 21 days of data (a total of 14 h) averaged together for each epoch. The error bars in wavelength are derived from the dispersion of the day averages. Errors in intensity are negligible. The maximum distance between the bisectors relative to the wavelength error occurs around an intensity level 20% down from the continuum. Here the formal error is 0.25 mÅ (12 m s^{-1}) while the separation is 1.5 mÅ (75 m s^{-1}).

Since May 1980 we have obtained full disk FTS bisector observations at 4–5-month intervals. As mentioned above, we have greater confidence in the FTS data than in those obtained with the spectrograph. For the FTS the error in wavelength is formally $\sim 0.04 \text{ mÅ}$ (2 m s^{-1}) based on six observations. The trend towards diminishing convection continues and is evident in the FTS bisector (Fig. 2b) even though solar maximum seems to have occurred in November 1979. Perhaps much of the flux which emerged at that time, largely concentrated in sunspots, is now slowly being distributed polewards to affect a growing fraction of the solar disk⁶.

These bisector studies are an outgrowth of our line-strength monitoring programme which began in 1975 and which has been pursued vigorously. The cooling trend indicated early on

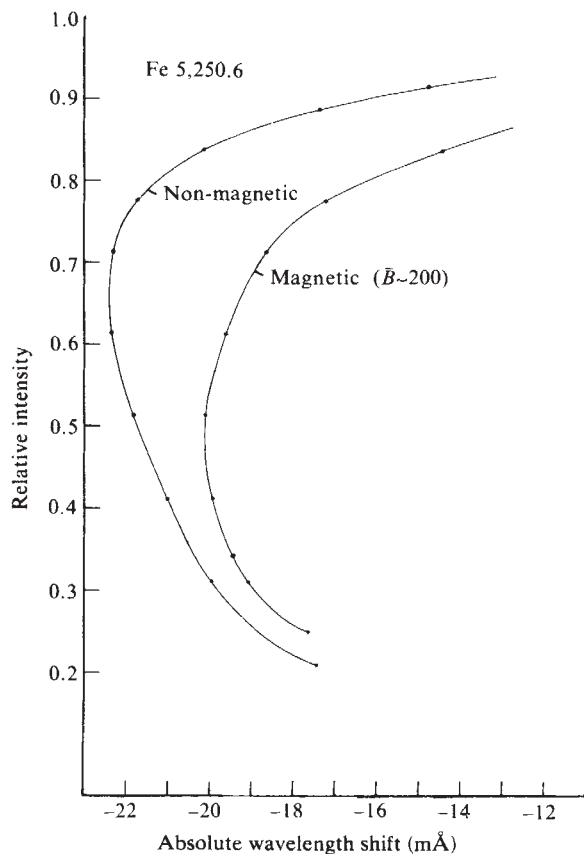


Fig. 1 Typical FTS line bisectors for adjacent magnetic and non-magnetic regions, observed at a spatial resolution of 1×2 arc min. The large, net blueshift arises from solar rotation. Note the non-magnetic C has larger amplitude and is more blueshifted than the magnetic C.

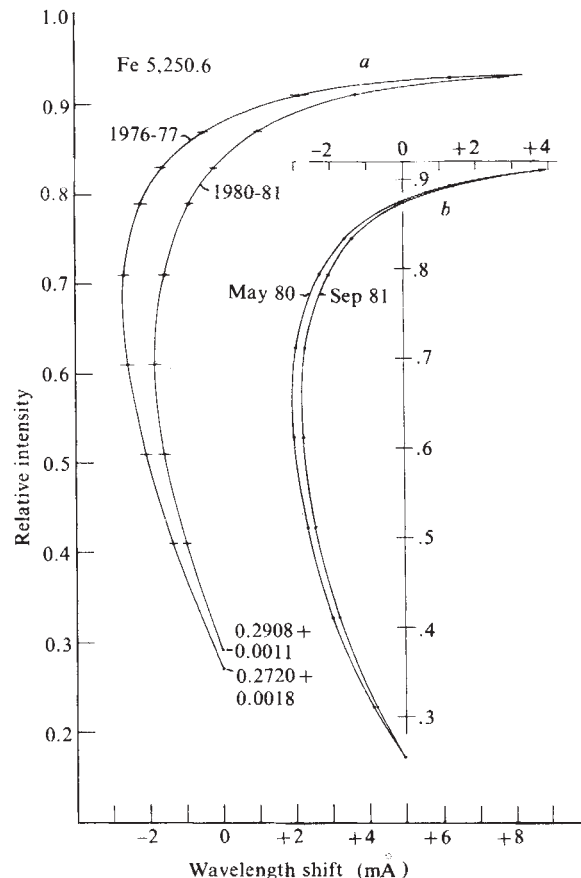


Fig. 2 a, Full disk spectrograph bisectors for solar minimum and maximum compared. The implication is decreased convection. b, As a except for a recent 16-month interval and using single day FTS data. In this case the formal errors are comparable with the size of the plotted points.

by a weakening of C 5,380 Å continues unabated. It was prematurely reported⁷ that Fe 5,379 Å, which has the opposite temperature response to C 5,380 Å, was synchronously strengthening, as would have been expected. Instead, 4 yr of additional data show it unequivocally to be weakening also. This apparent inconsistency was resolved by Holweger⁸ who used the fact that the two lines do not originate in the same atmospheric levels. He proposes that the low photosphere is temporally cooling, the upper photosphere is heating, and the result is a flattening of the temperature gradient. The quantitative implications of this change of temperature gradient on solar irradiance have not yet been determined.

I conclude that Fraunhofer lines in the Sun's integrated light spectrum display asymmetries which arise from convection. Observations resolving magnetic regions indicate that the amplitudes of the line asymmetries are smaller than in quiet regions. I have now observed a secular diminishing of the line asymmetry in full disk data from 1976 to 1981, corresponding to a period of increasing magnetic activity. Thus increasing surface magnetic flux (perhaps related to the phase of the activity cycle) is slightly retarding granular convection. Finally, this retarded convection may be the root cause of our observed line weakening and deduced solar cooling.

I thank Dr C. A. Lindsey for helpful suggestions.

Received 5 February; accepted 23 March 1982.

1. Spiegel, E. A. & Weiss, N. O. *Nature* **287**, 616–617 (1980).
2. Adam, M. G., Ibbetson, P. A. & Petford, A. D. *Mon. Not. R. astr. Soc.* **177**, 687–708 (1976).
3. Dravins, D., Lindgren, L. & Nordlund, Å. *Astr. Astrophys.* **96**, 345–364 (1981).
4. Nordlund, Å. *Astr. Astrophys.* (in the press).
5. Brault, J. W. *Ossini Mem. Oss. astrofis. Arcetri* **106**, 33–50 (1979).
6. Howard, R. & LaBonte, B. J. *Solar Phys.* **74**, 131–145 (1981).
7. Livingston, W. C. *Nature* **272**, 340–341 (1978).
8. Livingston, W. & Holweger, H. *Astrophys. J.* **252**, 375–385 (1982).

Constraints on the Moon's origin from the partitioning behaviour of tungsten

Horton E. Newsom* & Michael J. Drake†

Lunar and Planetary Laboratory, and Departments of *Geosciences and †Planetary Sciences, The University of Arizona, Tucson, Arizona 85721, USA

The Moon, Earth and eucritic meteorites display W/La ratios which correspond to depletions of W relative to chondrites by factors of ~19 (ref. 1). This depletion can be attributed to metal/silicate fractionation. Ordinary chondrites and CI chondrites with widely varying metal contents and oxidation states have similar W/La ratios, suggesting that nebular fractionations were unimportant. Therefore, assuming that both W and La are refractory and incompatible, the depletions of W relative to La in the Earth, Moon and eucrites are presumably due to separation of metal from silicate during planetary-scale igneous events. For the Moon, Rammensee and Wänke² concluded that any geophysically reasonable metallic lunar core would be too small to explain the large lunar depletion of W. It was therefore suggested² that the Moon acquired its depletion of W in a non-lunar igneous event, supporting the hypothesis that the Moon formed by fission from the proto-Earth, the upper portion of which was depleted in W due to terrestrial core formation³⁻⁵. We show here that this conclusion is implicitly based on a special case, namely equilibrium between metal and silicate phases with total melting of silicates. In the geologically more general case where metal fractionates from silicate at relatively low degrees of partial melting, the incompatible nature of W may negate this conclusion. A geophysically plausible metallic core can account for the depletion of W in lunar surface and, presumably, lunar mantle rocks. The low W/La ratio in the Moon cannot be used as unconditional evidence for a terrestrial origin of the Moon.

We first consider the initial concentrations of W and La in the Moon or in precursor lunar material. The presence of correlated pairs of refractory lithophile elements (for example, Ba/La) with CI ratios in the Earth, Moon and eucrites argues strongly for a chondritic W/La ratio in the Moon or precursor lunar material⁶. The abundances of W and La in lunar rocks and chondritic meteorites are plotted in Fig. 1. The magnitude of the lunar W depletion relative to CI and ordinary chondrites was calculated from the data in Fig. 1 and is listed in Table 1. The range of depletion factors for the Moon relative to the chondrites is 19 to 29, but there are large uncertainties in these values due to the uncertainties in both the lunar and chondrite W/La ratios.

We next consider the amount of metal that may be present in the Moon. Geophysical evidence for the metal content of the Moon is inconclusive. The tightest constraint on the size of a lunar core, by Hood *et al.*⁷, is a maximum radius ≤ 360 km, corresponding to 2% metal. Other investigations^{8,9} find indirect

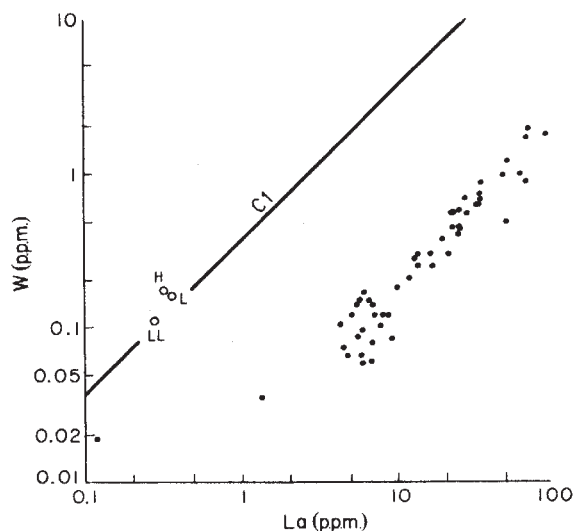


Fig. 1 Concentrations of W and La in lunar rocks, average values for three classes of chondrites, and the CI chondrite W/La ratio¹⁷⁻²⁰.

indications of a metal core with a radius of 200 to 500 km. A 500-km radius core corresponds to ~5.5% metal.

With information on the possible magnitudes of the lunar W depletion and the possible lunar metal content, we now evaluate the parameters which control the partitioning of W between metal and silicate in the Moon. The most important parameters are the degree of partial melting, because of the incompatible nature of W, and the metal/(silicate liquid) partition coefficient for W. The partition coefficient depends on the oxidation state, which depends on the FeO content, the Ni content of the metal and the temperature of the Moon. The bulk composition and, hence, oxidation state of the Moon is uncertain. The effects of these parameters on the W/La ratio are illustrated in Figs 2 and 3 and discussed below.

The behaviour of W at low degrees of partial melting is particularly important. Tungsten is an incompatible element, with siderophile affinities in lunar conditions of low oxygen fugacity. Therefore, W prefers metal to silicate melt and prefers silicate melt to silicate solid. At low degrees of partial melting the concentration of W in the silicate melt and in the metal is very high because W is substantially excluded from solid silicates. Partitioning of W between solid metal and liquid metal is unimportant because the partition coefficient is close to unity¹⁰. Figure 2 shows that, at low degrees of partial melting, significant depletion of W relative to La can occur even if the metal/(silicate liquid) partition coefficient $D(W)$ is relatively small. This is because the bulk metal/(silicate liquid + silicate solid) partition coefficient (bulk $D^{M/2S}$) is much larger than the $D(W)$ at low degrees of partial melting.

The effective degree of partial melting in the Moon during metal/silicate fractionation is important to determine because of its large effect on the bulk $D^{M/2S}$. In their calculations Rammensee and Wänke² implicitly assumed that total melting of the silicates occurred, and that $D(W)$ equals the bulk $D^{M/2S}$. In these conditions a very large $D(W)$ is required if a geophysically plausible metallic core is responsible for the depletion of W in the Moon. It seems geologically more reasonable, however, that equilibrium between silicate and metal be confined to small degrees of partial melting, because metal will rapidly sink and the melt will rapidly rise before a large degree of partial melting can occur. As the metal starts sinking and aggregating, the decreasing time scale and increasing length scale for diffusion will rapidly terminate equilibration of the silicates with the metal¹¹. A possible limit on the effective degree of partial melting is $F = 0.22$, where crystalline grains will no longer remain in contact with each other at the intergranular faces¹². The investigation by Walker *et al.*¹³ of the stability of

Table 1 W/La ratios in chondrites, and corresponding depletion factors for the Moon

Chondrite type	W/La ratios	Depletion factor (a) Moon/chondrite
CI	0.36 ± 0.04	19 ± 7
H	0.55 ± 0.06	29 ± 11
L	0.45 ± 0.11	24 ± 10
LL	0.40 ± 0.09	21 ± 9

Lunar W/La ratio = 0.019 ± 0.007 , data from Fig. 1.

the terrestrial low-velocity zone may also be relevant. They found that for $F=0.1$ the terrestrial low velocity zone could be largely drained of melt on a time scale of centuries, although this result is dependent on the scale of the melting system. Because of the large density difference between metal and silicate a value of $F < 0.1$ may be reasonable for the degree of partial melting during metal-silicate equilibration in the Moon, although the Moon may have continued to melt after effective isolation of the metal phase.

The other critical factor is $D(W)$, the metal/(silicate liquid) partition coefficient for the Moon. $D(W)$ depends strongly on the oxidation state of the Moon, which is related to the FeO content of the silicates and the Ni content of the metal in the Moon. The two scales plotted on the right ordinate of Fig. 2 show the relationship between $D(W)$ and the FeO content of the silicates for both 6% Ni and 52% Ni in the metal: 6% Ni corresponds to the assumption that the Ni/Fe ratio in the metal is approximately the cosmic ratio; 52% Ni assumes a chondritic model for the relationship between Ni content of the metal phase and total metal content¹⁴. The actual Ni content of metal in the lunar interior probably falls within these limits. Model lunar compositions cover a wide range in FeO content¹⁵ (2.3–13.9%), corresponding to a large range in $D(W)$. Most of the recent model compositions tend towards higher values of FeO corresponding to lower values of $D(W)$.

Figures 2 and 3 may be used to examine the conditions in which the depletion of W in the Moon could be explained by partitioning of W into metal within the Moon. The conditions discussed by Rammensee and Wänke² are a relatively high FeO content (13%), which corresponds to a low $D(W)$ of 53, and total melting of the silicates ($F=0.98$ in Fig. 3). For these

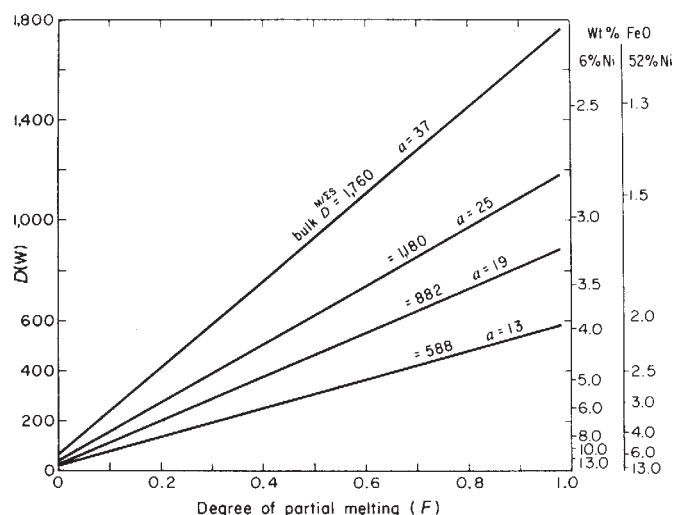


Fig. 2 The main constraints on the partitioning of W in the Moon constructed by calculating the bulk metal/(liquid silicate + solid silicate) partition coefficient (bulk $D^{M/LS}$), using an equilibrium partial melting model²¹, for different degrees of partial melting (F) and different metal/silicate liquid partition coefficients, $D(W)$. The solid silicate/liquid silicate partition coefficients used for both La and W are: olivine, 0.01; low-calcium pyroxene, 0.006; high-calcium pyroxene, 0.07; plagioclase, 0.14. Tungsten may be more incompatible than La, but no experimental mineral-melt data are known²². The model lunar silicate composition used was Hodges and Kushiro¹⁵. The diagonal lines represent the constant bulk $D^{M/LS}$ necessary to achieve a given depletion factor (a), assuming a metal content (x) of 2%. The relationship between bulk $D^{M/LS}$ and the depletion factor was calculated from the equation² $x = (a-1)/(bulk D^{M/LS} + a - 1)$. Breaks in slope occur when phases are exhausted, but the changes are small and are not shown. The two FeO scales plotted at the side were calculated by determining f_{O_2} (fugacity) for a given FeO content assuming 1,300 °C and ideal solution behaviour and 6% Ni or 52% Ni in the metal²³. The corresponding $D(W)$ was then calculated from the relation $\log D(W) = -\log f_{O_2} - 10.83$ determined experimentally by Rammensee and Wänke².

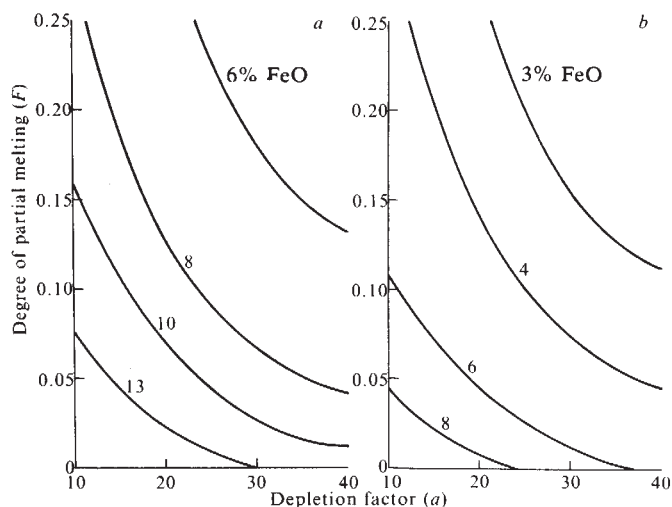


Fig. 3 *a, b*, The combination of per cent partial melting and FeO content required to achieve a reasonable depletion factor for the Moon (see Table 1): *a*, 2% metal containing 6% Ni; *b*, 2% metal containing 52% Ni. The plots were derived graphically from Fig. 2.

conditions Fig. 2 shows that a small metal content consistent with geophysical constraints could not be responsible for the observed depletion of W. We have argued, however, that equilibrium of silicate phases with metal at high degrees of partial melting is geologically unlikely. The situation at low degrees of partial melting is illustrated in a different way in Fig. 3*a* and *b*, which are derived directly from Fig. 2 for 6% Ni and for 52% Ni in the metal respectively. Figure 3*a* shows for the assumption of 6% Ni in the metal that reasonable lunar depletion factors (Table 1) can be achieved for lunar FeO contents as great as 13%. Figure 3*b* shows for the assumption of 52% Ni in the metal that reasonable depletion factors can be achieved only if the lunar FeO content is <6–8%. The actual ranges of values of $D(W)$ and F which are consistent with the lunar depletion of W are uncertain due to the wide range in possible lunar depletion values (Table 1) and the uncertainty in the FeO content of mafic silicates and Ni content of metal in the Moon. The only situation where the low W/La ratio in the Moon could not be explained by a 2% metal core is the combination of a high lunar FeO content with a very high Ni content in the metal. Even this constraint would be greatly relaxed if a core larger than 2% metal exists^{8,9}.

Of less importance, but of significance, is the variation in the modal composition of the Moon for different lunar bulk compositions. Compositions with low FeO contents tend to have high normative plagioclase contents. A large amount of plagioclase reduces the bulk $D^{M/LS}$ at low degrees of partial melting, because plagioclase may have a higher solid silicate/liquid silicate partition coefficient for W than does olivine or pyroxene.

A major question regarding depletion of W during partial melting is explaining the very homogeneous W/La ratios in lunar samples. Figure 3 shows the large sensitivity of the depletion factors to small changes in the degree of partial melting. Several factors may help explain this problem. A closer look at the actual mechanism of metal segregation is needed. Assuming that the metal grains are not all the same size, the partitioning of W and the settling of metal grains probably occur over a significant range in the degree of partial melting. This suggests that an important control on the depletion of W is the size-frequency distribution of metal. A relatively constant size-frequency distribution and metal content of accreting material may lead to a relatively constant W/La ratio. Other mechanisms following core formation could also lead to homogenization. These include convection and the possible formation of a lunar magma ocean.

We conclude from the above analysis that a metal content consistent with present geophysical constraints on the size of a

metallic lunar core could account for the depletion of W observed in lunar rocks. The low W/La ratio of the Moon cannot be used as evidence for the formation of the Moon from the Earth's mantle by fission following terrestrial core formation. The approximate coincidence of W/La ratios in the Earth, Moon and eucrite parent body may only reflect the operation of the same depletion mechanism in similar conditions on several parent bodies¹⁶.

We thank W. Rammensee for a helpful review. This work was supported by NASA grant NSG 7576.

Received 7 October 1981; accepted 23 March 1982.

1. Wänke, H. *et al.* *Proc. 4th lunar Sci. Conf.* 1461–1481 (1973).
2. Rammensee, W. & Wänke, H. *Proc. 8th lunar Sci. Conf.* 399–409 (1977).
3. Ringwood, A. E. *Origin of the Earth and Moon* (Springer, Berlin, 1979).
4. Delano, J. W. & Ringwood, A. E. *Proc. 9th lunar planet. Sci. Conf.* 111–159 (1978).
5. Wänke, H., Dreibus, G. & Palme, H. *Proc. 9th lunar Sci. Conf.* 83–110 (1978).
6. Dreibus, G., Kruse, H., Spettel, B. & Wänke, H. *Proc. 8th lunar Sci. Conf.* 211–227 (1977).
7. Hood, L. L., Sonett, C. P., Herbert, F., Cantrell, S. & Smith, L. *Lunar planet. Sci.* **12**, 457–459 (1981).
8. Russell, C. T., Coleman Jr, P. J. & Goldstein, B. E. *Lunar planet. Sci.* **12**, 914 (1981).
9. Stevenson, D. J. & Yoder, C. F. *Lunar planet. Sci.* **12**, 1043–1045 (1981).
10. Jones, J. & Drake, M. J. *Lunar planet. Sci.* **13**, 369–370 (1982).
11. Stevenson, D. J. *Science* **214**, 611–619 (1981).
12. Stocker, R. L. & Gordon, R. B. *J. geophys. Res.* **80**, 4828–4836 (1975).
13. Walker, D., Stolper, E. M. & Hays, J. F. *J. geophys. Res.* **83**, 6005–6013 (1978).
14. Anders, E. *Space Sci. Rev.* **3**, 583–714 (1964).
15. Warren, P. H. & Wasson, J. T. *Proc. 10th lunar Sci. Conf.* 2051–2083 (1979).
16. Newsom, H. E. & Drake, M. J. *Geochim. cosmochim. Acta* (submitted).
17. Rambaldi, E. R., Wänke, H. & Larimer, J. W. *Proc. 10th lunar planet. Sci. Conf.* 997–1010 (1979).
18. Palme, H. *et al.* *Proc. 9th lunar planet. Sci. Conf.* 25–57 (1978).
19. Mason, B. U.S. *geol. Surv. Prof. Pap.* 440-B-1 (1979).
20. Palme, H., Suess, H. E. & Zeh, H. D. in *Landolt-Börnstein Group 6, Vol. 2, Part a*, 252–257 (Springer, Berlin, 1981).
21. Consolmagno, G. J. & Drake, M. J. *Geochim. cosmochim. Acta* **41**, 1271–1282 (1977).
22. Palme, H. & Rammensee, W. *Proc. 12th lunar planet. Sci. Conf.* 949–964 (1981).
23. Myers, J. *EOS* **62**, 1070 (1981).

High-resolution coherent Raman spectroscopy of matrix-isolated carbon monoxide

I. R. Beattie, T. R. Gilson, S. N. Jenny & S. J. Williams

Department of Chemistry, The University, Southampton SO9 5NH, UK

IR spectroscopy of species isolated at low concentrations in inert matrices at cryogenic temperatures is an important physical technique in many branches of science. Raman spectra of such systems are much more difficult to obtain even at a resolution of only 1 cm⁻¹. Coherent Raman spectroscopy with an attainable resolution of at least 0.001 cm⁻¹ is capable of rivaling diode laser or Fourier transform IR spectrometers. We report here the first coherent Raman spectrum of a matrix-isolated species. The system chosen is carbon monoxide in nitrogen (15K) at 0.13% concentration with a resolution of 0.13 cm⁻¹.

Although coherent Raman spectroscopy is experimentally demanding compared with spontaneous Raman spectroscopy it can enable experiments to be carried out that would otherwise be impossible. Perhaps the most spectacular results are those of Owyong and co-workers¹ who used Raman loss spectroscopy (inverse Raman) to obtain spectra of ¹³CH₄ at 1.3 torr pressure. The resolution was Doppler limited at 0.009 cm⁻¹ resolution. A more familiar coherent Raman experiment is coherent anti-Stokes Raman spectroscopy (CARS) which is experimentally easier to set up (and less expensive) than a sophisticated Raman gain/loss experiment.

The selection rules for both Raman gain/loss spectroscopy and CARS are the same as for those of the spontaneous Raman effect². However there are important differences between the two coherent phenomena. Raman gain/loss spectroscopy is a two-photon process whereas CARS is a four-photon parametric process. The CARS intensity in phase-matched conditions is given by²:

$$I_3 = \left(\frac{8\pi^2 c^2 \omega_s}{h \omega_2^4} \right)^2 I_1^2 I_2 L^2 \sum_0 \Delta N_j^2 \left(\frac{d\sigma}{d\Omega} \right)^2 \left[\frac{1}{2\Delta\omega_j - i\Gamma_j} \right]^2$$

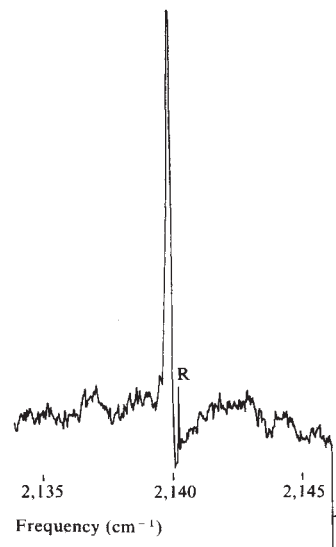


Fig. 1 The CARS spectrum of carbon monoxide in a nitrogen matrix at a concentration of 0.13%. (R = etalon reset.)

where ω_1 , ω_2 and ω_3 and I_1 , I_2 and I_3 are the pump (laser), Stokes (laser) and anti-Stokes (signal) frequencies and intensities respectively; ΔN is the population difference between ground and excited states; L is the interaction length; $d\sigma/d\Omega$ is the Raman scattering cross-section; $\Delta\omega_j$ is the Raman shift for the j th resonance; and Γ a convolution of natural and laser line widths. Note the squared dependence on ΔN , concentration, interaction length and $1/\Gamma_j$.

Consideration of the above factors led us to attempt to obtain CARS spectra of matrix-isolated species at cryogenic temperatures. Matrix isolation spectroscopy involves the trapping of a molecule (atom or fragment) in an inert substrate at low concentration so that the molecule is well isolated from other molecules. Conventional matrices are the inert gases and nitrogen. As guest molecules do not normally rotate in the host matrix, the lines will be sharp (frequently <1 cm⁻¹). Furthermore, at the isolation temperatures (of the order of 10K), the molecules will be in their ground vibrational states. Thus ΔN and $1/\Gamma_j$ are both maximized. For these experiments we chose nitrogen as the matrix gas because: (1) it gave us a strong signal from the substrate for initial lining up; (2) it has the advantage that the ¹⁴N¹⁵N (0.7 mol% concentration) peak may be used as a test for alignment and sensitivity. Carbon monoxide was chosen as the species to be isolated because not only has it been extensively studied, but also it has a vibrational frequency close to that of nitrogen.

Figure 1 shows the first coherent Raman spectrum of a matrix-isolated species. The concentration of the carbon monoxide is 0.13% and the resolution of the experiment is of the order of 0.13 cm⁻¹, which is an order of magnitude better than that obtained in a conventional matrix isolation Raman experiment. A further order of magnitude improvement in resolution could be obtained without significantly altering the nature of this experiment.

The spectrum was recorded using a carbon monoxide-nitrogen premix pulsed on to a calcium fluoride plate held in a Displex closed-cycle cooler operating in the region of 15K. The CARS system used a J K Lasers Q-switched Nd/YAG laser as pump, operating at 16.7 Hz repetition rate, with a 10 ns pulse length. The probe laser was a Chromatix CMX4 flash lamp pumped scanning dye laser amplified to ~100 kW peak power. For colourless matrices we find a laser damage threshold at a few hundred kilowatts peak power for a focused beam diameter of ~10⁻² cm. The spectra were recorded by scanning the CMX4 synchronously with (but in the opposite sense to) a Cary 81 double monochromator used as a high quality filter. The signal was detected by conventional means using a photo-

multiplier and boxcar detection. The frequency of the carbon monoxide transition was calibrated by running IR spectra on the same matrices, as the CARS spectra in these experiments were only recorded to an accuracy of $\pm 1 \text{ cm}^{-1}$.

Attempts to carry out studies in argon led to poor-quality spectra, which were only obtained at concentrations of the order of 1%. This is clearly due to line broadening coupled with interference between the resonant signal and the non-resonant background, leading to the observation of derivative type spectra. (There are several techniques available for minimizing this effect but they all lead to a reduction in signal.) In nitrogen the rotation of the carbon monoxide (which is thought to account for the line broadening in argon³) is presumably suppressed, leading to a narrow line.

Note that with Raman gain/loss spectroscopy a conventional Raman spectrum is obtained without the complications outlined above for CARS. There is also the possibility of a further order of magnitude improvement in resolution (leading to 0.001 cm^{-1} as a reasonable limit).

Received 31 December 1981; accepted 30 March 1982.

1. McDowell, R. S., Patterson, C. W. & Owyong, A. *J. chem. Phys.* **72**, 1071–1076 (1980).
2. Tolles, W. M., Nibler, J. W., McDonald, J. R. & Harvey, A. B. *Appl. Spec.* **31**, 253–272 (1977).
3. Dubost, H., Lecuyer, A. & Charneau, R. *Chem. phys. Lett.* **66**, 191–194 (1979).

Rapid localized glacio-isostatic uplift at Glen Roy, Scotland

J. B. Sissons & R. Cornish

Department of Geography, High School Yards,
Edinburgh EH1 1NR, UK

Many studies involving measurement of the altitudes of fossil marine and lacustrine shorelines have demonstrated glacio-isostatic tilting, particularly in North America, Fennoscandia and the British Isles. Such studies are normally illustrated by height–distance diagrams in which fossil shorelines are portrayed as straight lines. In some studies the shoreline–relation diagram is used to analyse and illustrate the altitudinal data^{1–3}; before constructing such a diagram it is assumed that a simple pattern of glacio-isostatic uplift has prevailed. Warping of shorelines, as opposed to uniform tilting, has been inferred in North America^{4–6}. On the other hand, Härmä⁷ envisaged block movements in Fennoscandia. Sauramo⁸ also proposed dislocations, but his interpretation was rejected by Hyyppä⁹. More recently the concept of the Scandinavian shield as a stable area has been dismissed^{10,11}. Here we provide evidence from Scotland for shoreline dislocation by differential movement of blocks of the Earth's crust.

The study of former crustal movements from fossil shorelines requires accurate, very closely spaced measurement, which has rarely been attempted outside Scotland. The results of such a study in the Forth Valley in central Scotland are shown in Fig. 1a. The upper shoreline, formed at 6,800 yr BP, is a surface feature while the lower shoreline, formed at 9,600 yr BP, is a buried feature¹². Both shorelines have corresponding sloping and essentially horizontal portions. Two parts of this area have been glacio-isostatically uplifted without significant tilting in the past 9,600 yr, while the remainder has been tilted. Furthermore, the older shoreline has two clear displacements, one of 1 m at the limit of the Loch Lomond (Younger Dryas) Advance, and one of nearly 1.5 m corresponding with a major fault: these displacements had ended by 6,800 yr BP (Fig. 1a).

In Glen Roy and vicinity in the Scottish Highlands, ice-dammed lakes existed during the Loch Lomond Stadial. We have recently levelled their fossil shorelines at 1,760 points.

Measurements were made at the break of slope representing the former shoreline (junction of 'parallel road' and backing cliff) and were spaced, where possible, at 25–30-m intervals. Traverses were normally 4–5 km long and the average closing error was 0.011 m. Where peat was present probes 1–3 m apart were aligned perpendicular to the shoreline. This work is more detailed and accurate than any previously undertaken on shorelines of former lakes.

We find that the shorelines are in part horizontal and in part tilted while, occasionally, there are displacements of <2 m. Despite these variations each shoreline has a limited altitudinal range. Using the means for groups of measured points and, where linear regression analysis was possible, the calculated values for the ends of tilted lengths of shoreline, the highest shoreline of Glen Roy varies between 349.5 and 351.9 m, the middle shoreline between 324.5 and 326.8 m, and the lowest one between 260.1 and 262.4 m, ranges of 2.4, 2.3 and 2.3 m respectively. There is a major exception to these generalizations which we consider here.

On the north-west side of upper Glen Roy a large landslide has destroyed all three 'parallel roads' (Fig. 2). Immediately south-west of the landslide all three shorelines rise above their normal altitude for this general area by ~3 m: the highest shoreline attains 354.1 m, the middle one 327.7 m and the lowest one 263.9 m. The 'roads' are here cut into a steep slope (25–30°) composed of metamorphic rocks of the Eilde Flags Formation (Moinian). Drift cover is thin and is mainly weathered bedrock. South-west from the landslide linear depressions roughly parallel with the contours mark the upper edges of slipped masses of bedrock (Fig. 2): they appear to be clear evidence of downhill movement but occur even where levelling proves the 'roads' to have been uplifted most.

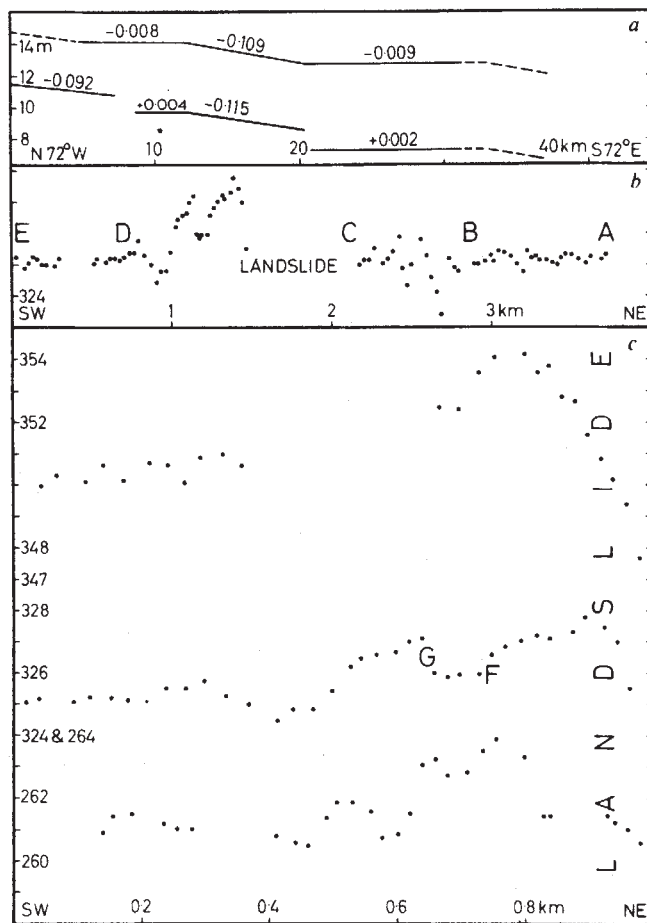


Fig. 1 Height–distance diagrams of dislocated shorelines. a, Western Forth Valley, gradients (in m per km). b, The middle shoreline of Glen Roy on both sides of the major landslide. c, All three Glen Roy shorelines south-west of the major landslide.

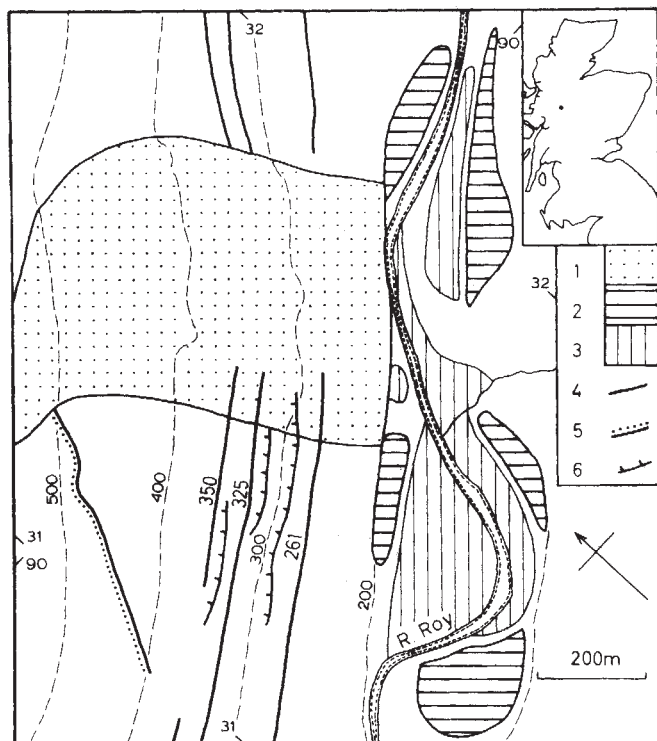


Fig. 2 The major landslide and other features. 1, Landslide; 2, part of the uniformly-sloping river terrace mentioned in the text; 3, other river terraces and floodplains; 4, 'parallel roads'; 5, inferred fault scarp; 6, upper edges of slipped masses of bedrock.

The middle shoreline, which is the best developed and sharply defined, is horizontal away from the major landslide: 16 levelled points in the south-west and 25 in the north-east (AB, DE, Fig. 1b) together average 325.1 m, the standard deviation being only 0.15 m. Between B and C the 'parallel road' is broken into several tilted portions: two of these rise to 0.7 m above shoreline altitude in the area of horizontality, indicating that BC is part of the uplifted area. South-west of the landslide the middle shoreline attains 2.6 m above its horizontal altitude. In the same locality the lowest shoreline attains 2.8 m above its altitude in the horizontal area and the highest shoreline 3.3 m above. All three 'roads' extend a short distance into the ground occupied by the major landslide, which has carried them down with it (Fig. 1c).

The point of maximal uplift on the highest shoreline lies upslope from the comparable point on the lowest shoreline. The culmination point of the middle shoreline is offset by 100 m: this difference is readily explained by the minor slips (for example, F and G, Fig. 1c) that have produced small but clear steps in the middle 'road'.

Localized uplift of the shorelines cannot be attributed to the landslide. Upward movement at the toe of a rotational landslide is well known: however, the Glen Roy feature is not a simple rotational slip and the toe is at the valley floor. Furthermore, the uplifted area extends hundreds of metres from the landslide. It is therefore most likely that crustal deformation occurred.

Pronounced uplift in a very small area suggests fracturing of the Earth's crust. Such fracturing is consistent with the presence of a linear bedrock ridge a few metres wide that slopes obliquely up the side of Glen Roy for a distance of 550 m (Fig. 2). Between the ridge and the valley side is a narrow depression that is up to 2 m deep despite being partly infilled by frost debris from the valley side and by peat. This feature is not a meltwater channel: the ridge is a clear positive feature standing above the adjacent ground (vertical aerial photographs reference O.S. 64:233, 032-033). Northwards the feature passes into the land-

slide. Southwards it begins to fade out before disappearing beneath frost debris that has moved downslope. Projected southwards the feature coincides exactly with the point where the middle shoreline begins to rise above its normal altitude (D, Fig. 1b). Part of the top 'road' is missing in the critical area and the bottom 'road' is here too poorly preserved to determine a relationship.

The linear nature of the ridge and associated depression and the evidence for uplift imply a fault. No detailed geological map of the area exists. Anderson¹³ mapped a straight fault 10 km to the NNE the projection of which coincides with the Glen Roy feature. The bend in the presumed fault corresponds with a small re-entrant in the hillside suggesting, along with the ridge, the presence of a reverse fault: this is consistent with crustal compression associated with glacio-isostatic uplift. Fault movement could have caused earthquakes, thus accounting for the large landslide.

The uplift clearly post-dates formation of the 'parallel roads'. The relative date of termination of uplift is also established. When the lake was emptied, probably very rapidly by *jökulhlaup*¹⁴, the new land surface encountered by the River Roy in the locally uplifted area was undulating, with large gravel fans alternating with lower ground. Initially the Roy cut into the fans and deposited the gravel thus obtained between them, producing a smooth profile now represented by a river terrace that extends intermittently along the whole length of the 2-km long uplifted area. The edge of the terrace is 110-170 m distant from the lowest shoreline. The standard deviation from a linear regression line ($r = 0.997$) of 41 levelled points on the terrace is 0.45 m. Hence pronounced differential uplift in this area had ceased by the time the terrace was formed. It is likely that the River Roy established this smooth profile shortly after lake drainage for the terrace is entirely in loose material (gravel) and, farther up the glen, was subsequently dissected by the Roy into a complex sequence of lower terraces. The major climatic amelioration that terminated the Loch Lomond Stadal occurred while the Roy ice-dammed lake stood at the highest shoreline. Hence the pronounced localized uplift occurred at the beginning of the Holocene.

Although minor earthquakes occur in Scotland today no related fault scarp has ever been detected. The country is essentially stable compared with the pronounced former instability implied by the Glen Roy evidence we have given. The latter is therefore attributed to glacio-isostatic uplift.

Our Glen Roy evidence thus shows that glacio-isostatic recovery can include abrupt localized uplift in an area now relatively stable. It adds to the Forth evidence (Fig. 1a) demonstrating differential movement of blocks of the Earth's crust. Because in both areas the dislocations post-date the Loch Lomond Advance, when glaciers were of limited extent, we suspect that larger movements occurred earlier during ice-sheet decay. The probability clearly exists that in Fennoscandia and North America, where ice sheets were far larger than in the British Isles, there are shoreline dislocations of even greater magnitude.

The NERC financed this research. Dr D. G. Sutherland criticized the draft of the paper.

Received 17 November 1981; accepted 26 March 1982.

1. Tanner, V. *Bull. Commn. géol. Finl.* **88**, 1-594 (1930).
2. von Post, L. *Geol. För. Stockh. Förh.* **69**, 293-320 (1947).
3. Sygne, F. M. & Stephens, N. *Trans. Inst. Br. Geogr.* **39**, 101-125 (1966).
4. Goldthwait, J. W. *J. Geol.* **16**, 459-476 (1908).
5. Leverett, F. & Taylor, F. B. *U.S. geol. Surv. Monogr.* **53**, 529 pp. (1915).
6. Broecker, W. S. *J. geophys. Res.* **71**, 4777-4783 (1966).
7. Härme, M. *Fennia* **89**, 29-31 (1964).
8. Sauramo, M. *Suomal. Tiedekat. Toim.* **III**, **51** (1958).
9. Hyypä, E. *Fennia* **89**, 37-48 (1964).
10. Lundqvist, J. & Lagerbäck, R. *Geol. För. Stockh. Förh.* **98**, 45-51 (1976).
11. Stephansson, O. & Carlsson, H. in *Earth Rheology, Isostasy and Eustasy* (ed. Möner, N.-A.) 327-337 (Wiley, Chichester, 1980).
12. Sissons, J. B. *Trans. Inst. Br. Geogr.* **55**, 145-159 (1972).
13. Anderson, J. G. C. *Trans. R. Soc. Edinb.* **63**, 15-36 (1956).
14. Sissons, J. B. *J. geol. Soc. Lond.* **136**, 215-224 (1979).

Exotic terranes of western California

M. O. McWilliams

Geophysics Department, Stanford University, Stanford, California 94305, USA

D. G. Howell

US Geological Survey, Menlo Park, California 94025, USA

Numerous distinct geological terranes compose the North American Cordillera¹; there may be as many as 50 terranes in California alone². Critical to deciphering the history of Cordilleran tectonic assembly is an understanding of the displacement history of individual terranes. It is therefore important to know: (1) whether a terrane has undergone significant motion with respect to the stable craton (that is, whether it is allochthonous or exotic); (2) if so, when relative motion started and stopped; (3) from where an individual terrane originated; and (4) the nature of inter-terrane movements. We consider here the problem of determining whether the now-juxtaposed Salinian and Stanley Mountain terranes of California became amalgamated at or near their present position with respect to cratonic North America, or if they collided at a considerable distance from their present positions and were later accreted to North America as a composite package. The palaeomagnetic data that we present indicate that the latter was the case.

Terranes are fault-bounded geological entities of regional extent, each characterized by a geological history that is different from that of adjacent terranes. Such geological histories are determined from stratigraphical successions preserved within terranes, but in some cases this information is obscured owing to tectonic or sedimentological disruption or metamorphic overprinting. Composite terranes are produced by the amalgamation of two or more terranes. The southern Coast Ranges of California west of the San Andreas Fault consist of two composite tectonostratigraphical packages called the Salinian and Sur-Obispo terranes (Fig. 1). Basement rocks of the Salinian terrane are Middle and Upper Cretaceous granitic plutons and older high-temperature metasedimentary rocks bounded on the west by the Sur-Nacimiento Fault. To the west lies the Sur-Obispo composite terrane, represented by the Stanley Mountain and San Simeon terranes. The Stanley Mountain terrane consists of Middle Jurassic ophiolite and an overlying thick sequence of Upper Jurassic to Upper Cretaceous forearc-basin strata. Structurally below the Stanley Mountain terrane is Franciscan mélangé of the San Simeon terrane. The Stanley Mountain strata closely resemble, and are coeval with, the Great Valley sequence³, which is typified by exposures along the west sides of the Sacramento and San Joaquin Valleys. Truncation by faults and unconformities disrupts the section in most places, but a nearly complete section of 4,000 m of Upper Jurassic and Cretaceous strata is exposed in the southeastern San Rafael Mountains⁴. These forearc-basin strata are demonstrably allochthonous with respect to Salinian basement, yet in late Cretaceous time the two terranes seemingly collided to form a new amalgamated terrane, as Campanian and younger sedimentary rocks of both terranes are petrographically identical and are characterized by non-marine horizons of locally derived granitic detritus^{5,6} (Fig. 2). Several coherent packages of Upper Cretaceous arkosic strata in the San Simeon terrane may indicate continuity with the Salinian and Stanley Mountain terranes; however, equivocal sedimentary sequences which lap across these terranes are no older than early Tertiary.

Commencing in Campanian time with the paralic granite-grus lithofacies, the forearc took on a borderland-like configuration with detritus spilling into contrasting depositional settings on the newly juxtaposed terranes^{6,7}. An episode of wrench faulting accompanied the suturing presumably along a now-distant margin of the Cordillera because Upper Cretaceous and Lower Tertiary strata contain numerous clasts of siliceous vol-

canic porphyries admixed with the granitic debris, yet such a volcanic provenance does not exist on either the Salinian or Stanley Mountain terranes.

Palaeomagnetic inclination data from Upper Cretaceous strata of the Salinian terrane indicate that deposition occurred as much as 2,500 km south of their present location (ref. 8; M. E. Williams, M. Schnapp and A. Cox, in preparation; D. E. Champion, personal communication). We report here new data which indicate a similar exotic origin for the Jurassic and Cretaceous rocks of the Stanley Mountain terrane.

To determine the palaeolatitude of the Stanley Mountain terrane in Jurassic time, samples of pillow basalt were collected at 10 sites spanning an 800 m section of ophiolite at Stanley Mountain (Figs 1, 2). Alternating field (a.f.) demagnetization of pilot specimens from each site yielded stable, well grouped magnetization directions at eight sites after removal of young secondary magnetizations of probable recent viscous origin; one example is shown in Fig. 3a. After partial demagnetization in peak alternating fields of 60–80 mT, the eight mean directions are well grouped and define an overall *in situ* mean direction $D = 053^\circ$, $I = +54^\circ$, $\alpha_{95} = 8^\circ$. This direction is distinct from the present geomagnetic and axial dipole field directions for the sampling locality, as well as from any expected field direction for stable North America since early Mesozoic time⁹. A reasonable explanation for the *in situ* direction is that it represents a sampling of a primary Jurassic magnetization (TRM or high-temperature CRM) and should therefore be restored to palaeohorizontal attitude, thereby obtaining the palaeolatitude at which the rocks acquired their magnetization. If the magnetization is primary, the overall tectonically corrected mean direction is $D = 041^\circ$, $I = -27^\circ$, $\alpha_{95} = 8^\circ$, corresponding to a palaeolatitude of $14 \pm 7^\circ$. Structural attitudes of sedimentary rocks which immediately overlie the basalts were used as a palaeohorizontal reference indicator. If the age of the pillow basalt is 160 ± 5 Myr, as indicated by isotopic data from other Coast Range ophiolite samples¹⁰, then a northerly or southerly

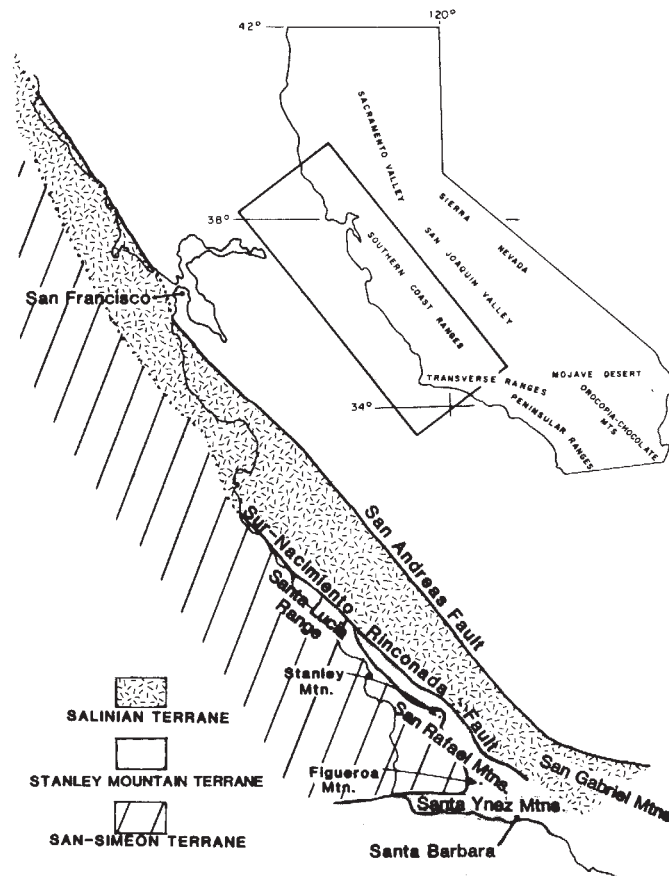


Fig. 1 Geographical sketch map of Stanley Mountain, San Simeon and Salinian terranes in western California.

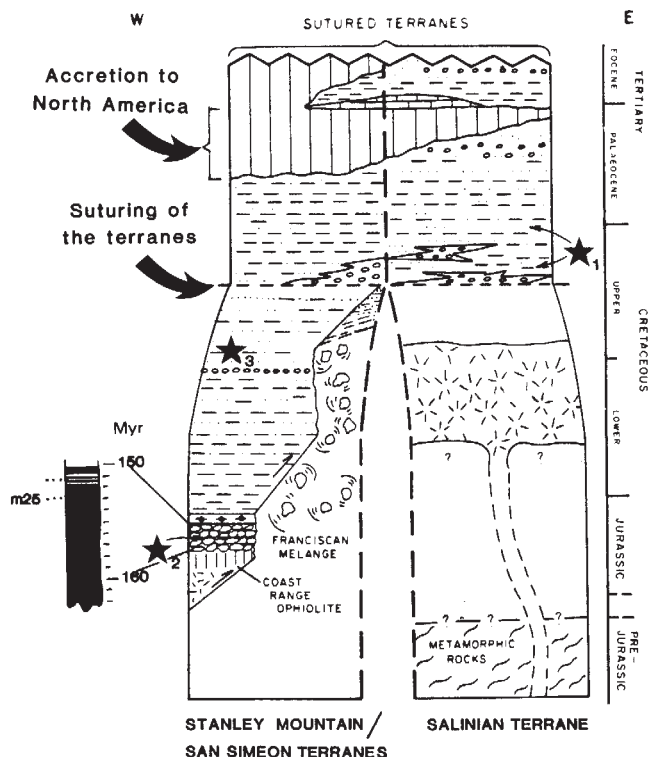


Fig. 2 Generalized stratigraphical columns for Salinian and Stanley Mountain/San Simeon terranes. Suturing of the two terranes was probably completed by the late Cretaceous, as equivalent strata lap onto both terranes after this time. ★, Stratigraphical positions of palaeomagnetic studies (a, D. E. Champion, personal communication; b, ref. 8 and M. E. Wilson, M. Schnapp and A. Cox, in preparation). The inset on the left is magnetic reversal time scale¹¹ (black = normal polarity).

palaeolatitude cannot be invoked, for the geomagnetic field was of mixed polarity at this time¹¹ (see reversal time scale inset, Fig. 2).

If the magnetization were of secondary origin and dated from Tertiary time, an exact tectonic correction would not be possible because the structural attitude at the time of remagnetization would be unknown. One possibility would be that the basalts acquired their magnetization after tectonic tilting and were subsequently rotated 50–60° clockwise about a vertical axis. At the locality where the samples were collected, the basalts form part of a steeply dipping homoclinal sequence, so that a fold test was not possible. There is no evidence, however, for Tertiary remagnetization by a thermal mechanism, for isotopic ages are not reset and reheating effects are not visible in hand specimen or thin section.

The sedimentary record indicates that the Stanley Mountain and Salinian terranes probably collided in the late Cretaceous. To determine the palaeolatitude of the Stanley Mountain terrane at approximately the time of collision, samples of Upper Cretaceous (Cenomanian–Santonian) flysch strata were collected at 10 sites in the San Rafael Mountains near Figueroa Mountain (Figs 1, 2). After a.f. demagnetization in peak fields up to 80 mT, a single stable component of magnetization was isolated (an example is shown in Fig. 3b). The overall *in situ* mean direction from these 10 sites is $D = 002^\circ$, $I = +74^\circ$, $\alpha_{95} = 10^\circ$, which is significantly different from both the present geomagnetic field direction and the axial dipole field direction at the sampling locality, suggesting that it is not of recent origin. Also, this overall mean direction is significantly different from any geomagnetic field direction since early Mesozoic time derived from stable North America⁹. On tectonic correction to palaeohorizontal, the overall mean direction is $D = 023^\circ$, $I = +12^\circ$, $\alpha_{95} = 11^\circ$. The minor decrease in overall precision accompanying this correction is not statistically significant and may be due to pre-magnetization slumping as noted in other

palaeomagnetic studies of turbidites⁸. The slightly skewed distribution of site mean directions may be due to our inability to make exact tectonic corrections which compensate for this post-depositional slumping.

If the magnetization of these samples is primary and dates from the time of deposition or soon thereafter, the tectonically corrected inclination of $12 \pm 10^\circ$ would correspond to a palaeolatitude of $6 \pm 5^\circ$ (N or S). This range is slightly lower than the range of apparent palaeolatitudes from the Pigeon Point Formation on the Salinian terrane (ref. 8, and M. E. Wilson, M. Schnapp and A. Cox, in preparation), but if a Northern Hemisphere interpretation is invoked (Fig. 4), this might be expected, as the Pigeon Point strata are slightly younger (Campanian–Maestrician). An alternative explanation for the palaeolatitude discrepancy could be that the suturing of the two terranes might not have been completed by this time.

The palaeomagnetic data shows that (1) the Salinian composite terrane lay in low ($21 \pm 7^\circ$) palaeolatitudes⁸ in Campanian–Maestrician time, and occupied higher ($25 \pm 3^\circ$) palaeolatitudes¹⁰ in early Palaeocene time; (2) the Stanley Mountain terrane lay in slightly lower ($6^\circ \pm 5^\circ$) palaeolatitudes in Cenomanian–Santonian time, and occupied slightly higher ($14^\circ \pm 7^\circ$) palaeolatitudes at about 160 ± 5 Myr; either a north or south palaeolatitude is possible; (3) a simple displacement model (solid line in Fig. 4) would place the Sur-Obispo composite terrane at 14° S at 160 Myr, at 6° N at about 90 Myr and at 21° N at about 70 Myr, after colliding with the Salinian composite terrane; and (4) alternatively, the Stanley Mountain terrane could have occupied low northerly palaeolatitudes from Jurassic to Cretaceous time, beginning its major northward migration at about the time it collided with the Salinian terrane (dashed line in Fig. 4). Coupled with the geological arguments, these data indicate that the late Cretaceous collision of the two terranes occurred in low palaeolatitudes, far from the present California continental margin.

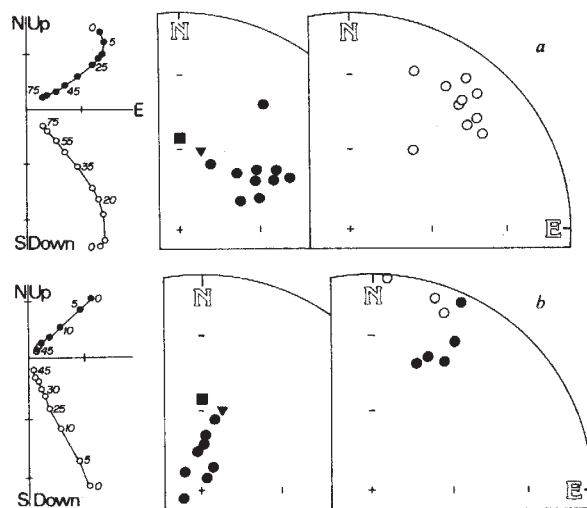


Fig. 3 a, Palaeomagnetic results from Stanley Mountain pillow basalts. Left: orthogonal vector projection of a.f. demagnetization experiment; circles represent projection of magnetization vector on the horizontal (●) or east-west vertical (○) plane; axis ticks $0.2 \times 10^{-3} \text{ A m}^{-1}$. Numbers are peak a.f. treatment in units of 10^{-3} T . Centre: equal area projection of site mean magnetization directions after a.f. demagnetization. No tectonic correction is applied; each direction plotted is the average of six individually oriented samples. Solid symbols represent positive (downward) inclinations. ■ and ▼ are axial dipole field and present geomagnetic field directions respectively. Right: as in centre, but with magnetization directions restored to palaeohorizontal. Open symbols represent negative (upward) inclination. b, Palaeomagnetic results, Upper Cretaceous sedimentary rocks from Figueroa Mountain. Left: as above, axis ticks $0.1 \times 10^{-3} \text{ A m}^{-1}$. Centre: site mean directions before tectonic correction; conventions as above. Right: site mean directions before tectonic correction; conventions as above.

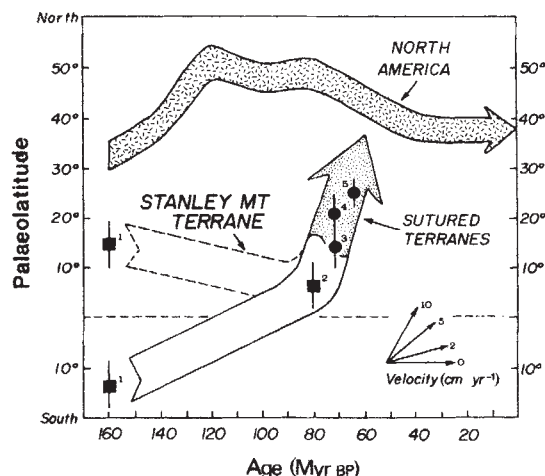


Fig. 4 Palaeolatitude of the Salinian and Stanley Mountain terranes as a function of time, compared with palaeolatitude of 'target area' on North America for present-day position, 35–39° north latitude. Stippled path represents the two terranes after suturing. The North American and Salinian/Stanley Mountain tracks intersect at ~50–55 Myr BP to allow for Eocene reconstruction of the Butano-Point of Rocks submarine fan system. ■, Data from Sur-Obispo terrane; ●, data from Salinian terrane. (1) and (2) are from this study; (4) from ref. 8; (3) from M. E. Wilson, M. Schnapp and A. Cox, in preparation; and (5) D. E. Champion, personal communication. Arrows in lower right give northward velocity in cm yr^{-1} for reference. The dashed and solid lines represent alternative palaeolatitudes for the Stanley Mountain terrane depending on the choice of magnetic polarity.

The classic sequence of events proposed for the evolution of the western Cordillera involves three general episodes^{12–14}: (1) a late Jurassic to early Tertiary subduction regime with major deposition in an arc-trench gap configuration; (2) early Tertiary proto-San Andreas faulting and a borderland setting during an early to middle Tertiary episode of oblique subduction along a flattened Benioff-Wadati zone; and (3) continental margin transform faulting resulting in tectonic splintering and additional borderland development. If recent palaeomagnetic studies (ref. 8; M. E. Wilson, M. Schnapp and A. Cox, in preparation; D. E. Champion, personal communication), including our own, are accurate and basin analyses^{4,6,7} correct, both the timing of events and nature of origin of the continental margin must be reinterpreted.

Analysis of the 1,500 m thick fluvio-deltaic sequence in the southeastern Stanley Mountain terrane indicates that the Stanley Mountain and Salinian terranes were probably attached by ~72 Myr BP. Thus the Stanley Mountain and Salinian terranes, both of which come from quite different parent terranes, must have travelled northwards together, despite the remarkable parallelism of Cretaceous depositional environments between the Stanley Mountain terrane and the Great Valley sequence of the Sacramento and San Joaquin Valleys. However, the Stanley Mountain/Salinian composite package was probably emplaced in California by the early Eocene, based on the inference that submarine fans sequences of this age¹⁵ lap across these terranes and other terranes autochthonous to California.

The palaeomagnetic data presented here and elsewhere indicate a large northward component of relative plate motion along the western margin of North America during the Cretaceous and early Tertiary. Several late Mesozoic terranes presumably originated at the latitude of present-day Central America and moved northwards to the California area by late Palaeocene time. Plate motion reconstructions (ref. 16 and D. C. Engebretson, in preparation) demonstrate that Farallon-North America relative motion for most of the Cretaceous would have been largely orthogonal to a north-west trending margin of western North America. Kula plate motion relative to North America had a significant northward component at high latitudes, but owing to an Euler pole position in north-

western South America, Kula-North America relative motion in the Central American regions would also have been largely orthogonal (ref. 17 and D. C. Engebretson, in preparation).

Two possible ways to account for the northward shifting of terranes are immediately apparent¹⁷: (1) reconstruct the late Mesozoic Pacific basin plates including now vanished plate(s) whose configurations would account for the northward slip required by the palaeomagnetic data and for the episodic plutonism mapped all along the Cordillera, and/or (2) palinspastically restore Central America and the ancient Caribbean plate westwards to increase the distance from the Kula-North America Euler pole in late Mesozoic time, so that Kula plate motion could better account for the northward component of motion of the Stanley Mountain and Salinian terranes.

We thank Anne Demitrack, Brian Hausback, Bill Keller, Peter Lellis and Norman Sleep for assistance in sample collection, David Engebretson and Ben Page for editorial comments, and especially Michel Debiche and Anne Demitrack for many laboratory measurements. The research was supported by grant PRF-11405 from the Petroleum Research Fund of the American Chemical Society.

Received 21 January; accepted 9 March 1982.

1. Coney, P. J., Jones, D. L. & Monger, J. W. H. *Nature* **288**, 329–331 (1980).
2. Blake, M. C., Howell, D. G. & Jones, D. L. *U.S. geol. Surv. Open File Map Ser.* (in the press).
3. Page, B. M., in *The Geotectonic Development of California* (ed. Ernst, W. G.) (Prentice-Hall, Englewood Cliffs, New Jersey, 1981).
4. Howell, D. G. et al. in *Soc. Economic Paleontologists and Mineralogists, (Pacific Sec.), Pacific Coast Paleogeogr. Field Guide Vol. 2* (eds Howell, D. G., Vedder, J. H. & McDougall, K.) 1–46 (1977).
5. Howell, D. G., McLean, H. & Vedder, J. H. *EOS* **61**, 949 (1980).
6. Vedder, J. H., Howell, D. G. & McLean, H. *Mem. Am. Ass. petrol. Geol.* (in the press).
7. Howell, D. G. & Vedder, J. H. *Soc. Economic Paleontologists and Mineralogists, (Pacific Sec.), Pacific Coast Paleogeography Symp. Vol. 2* (eds Howell, D. G. & McDougall, K. A.) 107–116 (1978).
8. Champion, D. E., Gromme, C. S. & Howell, D. G. *Soc. Economic Paleontologists and Mineralogists, (Pacific Sec.), Field Trip Guide, Upper Cretaceous and Paleocene Turbidites, Central California Coast*, 53–56 (1981).
9. Irving, E. *Can. J. Earth Sci.* **16**, 669–694 (1979).
10. Hopson, C. A., Mattinson, J. M. & Pessagno, E. A. in *The Geotectonic Evolution of California* (ed. Ernst, W. G.) (Prentice-Hall, Englewood Cliffs, New Jersey, 1981).
11. Cox, A. in *The Geologic Timescale* (ed. Harland, W. B.) (Cambridge University Press, in the press).
12. Dickinson, W. R. *Rev. geophys. Space Phys.* **8**, 813–860 (1970).
13. Dickinson, W. R. *Am. J. Sci.* **272**, 551–576 (1972).
14. Dickinson, W. R. & Rich, E. I. *Bull. geol. Soc. Am.* **83**, 3007–3024 (1972).
15. Nilsen, T. H. & Clarke, S. H. *U.S. geol. Surv. Professional Pap.* **925**, 1–64 (1975).
16. Morgan, W. J. in *The Sea* (ed. Emiliani, C.) (Wiley, New York, 1980).
17. Page, B. M. *Am. J. Sci.* (in the press).

Implications of Palaeozoic phosphorites in the northern Sierra Nevada range

Robert J. Varga

Union Oil Research Center, PO Box 76, Brea, California 92621, USA

Marine phosphorites are known to occur within multiply deformed Ordovician/Silurian and Upper Devonian rocks of the northern Sierra Nevada range, California. Although the phosphatic sediments constitute a small percentage of total lithologies in the area, they provide unique information regarding the age and palaeogeographical setting of these rocks. The mode of occurrence of the Sierran phosphorites and their association with black shale and chert suggest that they may have formed along a continental margin which experienced strong equatorial or trade wind belt upwelling. Consideration of worldwide plate reconstructions for the early to middle Palaeozoic shows that the present western margin of North America was one of the few favourable sites for phosphate formation during this time period. Thus, the presence of marine phosphorites within Palaeozoic rocks of the northern Sierra Nevada is consistent with arguments which suggest that these rocks were deposited near the western margin of North America. Although considerable margin-parallel translations cannot be ruled out for these rocks, they do not appear to constitute a far-travelled microplate derived from a distant, trans-oceanic site.

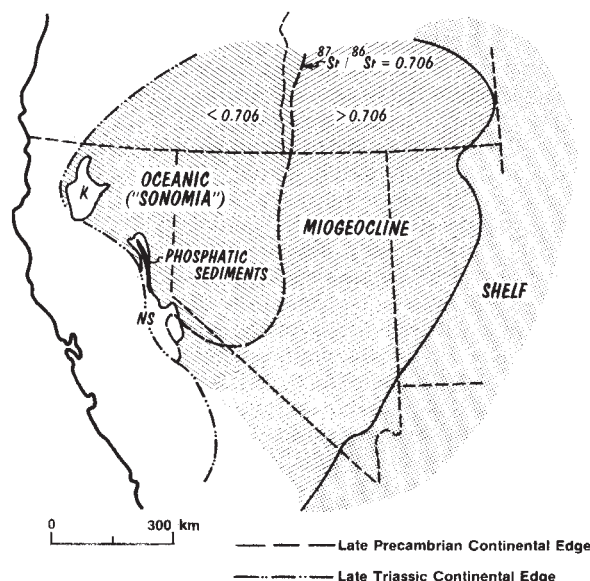


Fig. 1 Sketch map showing lower to middle Palaeozoic petro-tectonic assemblages of the US Cordillera. Generalized outcrops of Palaeozoic strata in the northern Sierra Nevada (NS) and eastern Klamath Mountains (K) are shown with no pattern within 'oceanic' assemblage. Data from refs 19, 38, 44.

Phosphorite occurs within the pre-late Devonian (in part Ordovician/Silurian) Shoo Fly Complex (SFC) and late Devonian Elwell and Sierra Buttes Formations (EF and SBF, respectively) of the northern Sierra Nevada range. These formations and their phosphatic components crop out in an elongate, north-west-trending belt more than 150 km in length (Fig. 1). The SFC comprises predominantly siliciclastic sediments, in part of continental derivation¹, as well as minor limestone, dolomite and mélange^{2,3}. Black shale, chert and intermediate to silicic volcanics of the SBF and EF⁴ overlie the SFC with angular discordance⁵⁻⁷. All three formations are multiply deformed⁷⁻⁹ and were metamorphosed to lower greenschist facies at ~150 Myr (ref. 10). The SBF and EF constitute components of a middle Palaeozoic to early Triassic volcanic arc terrane

which extends to the eastern Klamath Mountains of northern California¹¹.

Metamorphism and intense deformation has hindered interpretation of the depositional environments of Palaeozoic strata in the northern Sierra Nevada. As discussed below, such information is critical to the evaluation of their pre-Mesozoic palaeogeographical position. Thus, the presence of a unique sedimentary feature, such as phosphorite, assumes heightened significance. It is, therefore, of interest to compare the Sierran phosphorites with well-documented occurrences of marine phosphates about which some palaeoenvironmental generalizations can be made.

Phosphate within Palaeozoic rocks of the northern Sierra Nevada occurs principally as nodules, and to a lesser degree as lenses and angular fragments within foliated black shale and chert. Phosphate nodules (Fig. 2a) are generally round to sub-round and vary in size from ~0.5 to 2.5 cm. The nodules are black on fresh surface and weather to a light grey to white colour. The porous nodule surfaces often show a crude concentric pattern with inner layers composed of darker, more porous material than outer layers. Phosphatic lenses commonly accompany phosphate nodules (Fig. 2b). These lenses are up to 2.5 cm thick and 1 m in length. Angular fragments of phosphate, up to 0.5 cm in length, are ubiquitous associates of both phosphate nodules and lenses.

Phosphate nodules are composed of dark brown, cryptocrystalline, isotropic collophane (Fig. 2c) with inclusions of silica and pyrite. Slight differences in colour, transparency and quartz content define a crude, discontinuous banding towards the outer boundaries of most nodules. Every nodule examined contains a characteristic black, iron-rich outer surface which is up to 0.6 mm wide (Fig. 2c). Silica occurs as fine-grained intergrowths within the collophane matrix, as irregular masses up to 0.2 mm in size, and as recognizable biogenic remains (Fig. 2d). Well-preserved sponge spicules and radiolaria have been recovered by HCl dissolution of nodules from the EF¹² and SFC⁷.

The average of four microprobe analyses of nodule interiors from the SFC is as follows: $P_2O_5 = 39.96\%$, $F = 3.56\%$, $CaO = 52.04\%$, $SiO_2 = 1.71\%$, $Cl = Sr = Na_2O = SO_3 = MgO = K_2O < 500$ p.p.m., total = 97.27% (the slightly low total could not be accounted for and presumably represents a small amount of structural OH^- , instrument calibration error, or both). CaO/P_2O_5 and F/P_2O_5 ratios in the SFC nodules have values of 1.30 and 0.09, respectively, which are close to the ratios

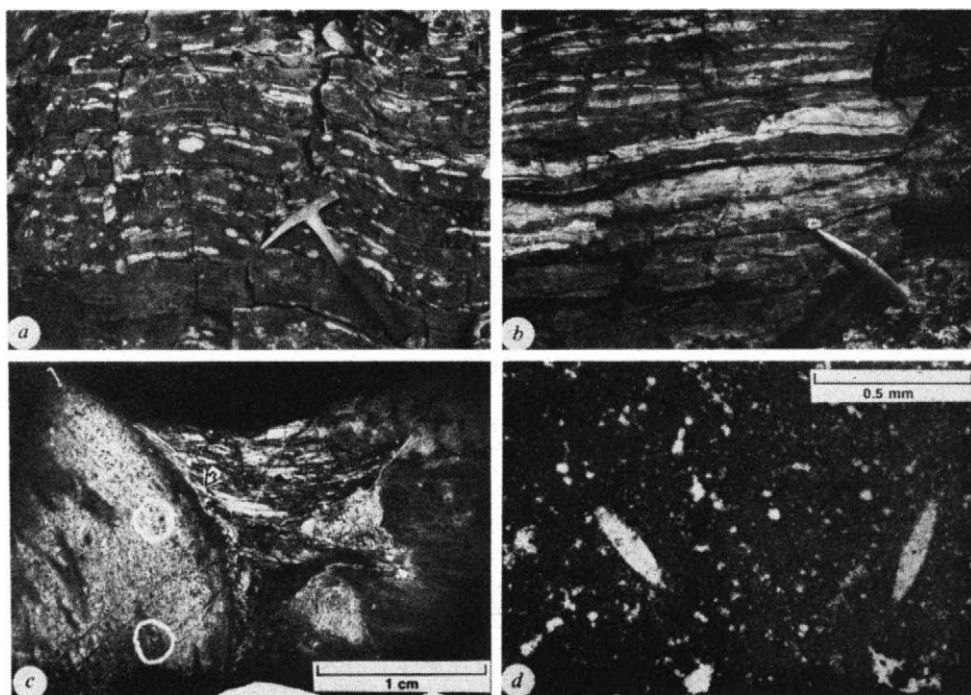


Fig. 2 a, Phosphate nodules (white areas) within black chert of the Elwell Formation. b, Phosphate lenses (white areas) and nodules (near hammer tip) within Shoo Fly Complex. c, Negative photograph of thin section showing the boundary between an undeformed phosphate nodule from the SFC (left side of photograph) and highly deformed black shale matrix. Note faint concentric banding within nodule and dark, iron-rich coating (light band in photo) on outer nodule surface (tip of arrow). Circled areas are probe sites. d, Sponge spicules(?) within SFC nodule.

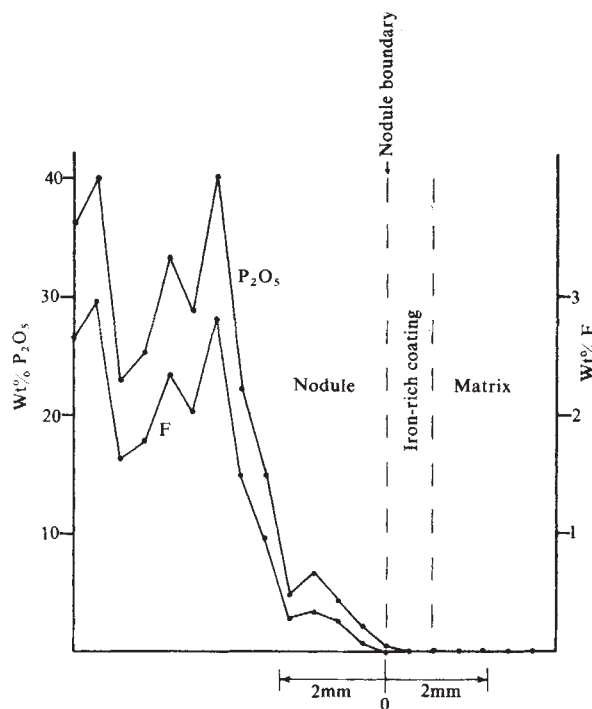


Fig. 3 Microprobe data from a 9.5-mm traverse across an SFC nodule-matrix boundary. Peaks and valleys in P_2O_5 and F contents show the relative abundances of apatite within the nodule. Areas of high apatite concentration correspond to layers of low included silica content. Iron-rich coating is seen as the dark band in Fig. 2c which separates matrix from nodule.

expected (1.32, 0.09) in unsubstituted fluorapatite, $Ca_{10}(PO_4)_6F_2$. CO_2 content was estimated using Gulbrandson's¹³ X-ray peak-pair method. By this method, CO_2 content of two SFC nodules and one EF nodule average $0.38 \pm 0.56\%$ and $0.74 \pm 0.56\%$, respectively. These estimates are consistent with a nearly ideal fluorapatite composition of the Sierran nodules.

The composition of the Sierran nodules is unusual in that most marine apatites contain 3–6% CO_3^{2-} substituted into the PO_4^{3-} site with the charge balanced by a variety of ions¹⁴. Post-depositional weathering, as a means of removing CO_2 and other components from the apatite structure¹⁵, seems an unlikely mechanism to explain the apatite chemistry of the Sierran nodules. Outcrops from which the samples were collected for this study were quite fresh, having been scoured by alpine glaciers during the Pleistocene. Thus, the present chemistry of the nodules may represent either their original chemistry or the effects of chemical exchange during Jurassic dynamothermal metamorphism.

The origin of the concentric banding in Sierran phosphate nodules (Fig. 2c) is of interest because of the observation that many recent phosphorites contain films and layers of manganese¹⁶. Banding within an SFC nodule was investigated optically and by a 9.5-mm microprobe scan (Fig. 3) for P, F and Mn from the interior of the nodule to its outer edge. Weight percentages of both P_2O_5 and F, and thus apatite, show peak concentrations 3.3 and 5.6 mm from the outer boundary of the nodule with a drop in concentration in between. From 3.3 mm to the nodule boundary the concentration of apatite steadily decreases to zero with effectively no phosphorous or fluorine in the matrix. Mn was not present above detection limits (500 p.p.m.), anywhere along the microprobe traverse. X-ray fluorescence of the dark outer coating surrounding the nodule (Fig. 2c) showed it to be iron-rich with small amounts of titanium, calcium and potassium. Banding within nodule interiors thus corresponds to differences in apatite concentration. It is apparent optically that an increase in intergrown silica generally corresponds to decreases in apatite content. The sharp

decrease in apatite content towards the nodule boundary is interpreted as indicating a steady drop in accretion of apatite during the waning stages of nodule growth.

The pre-Mesozoic position of Palaeozoic rocks of the northern Sierra Nevada with respect to North America has caused some debate¹⁷. The question is whether these rocks represent autochthonous or parautochthonous terranes which have been associated with the western margin of North America since the early Palaeozoic^{7,11,18} or whether they represent components of a far-travelled island arc terrane, Sonomia (Fig. 1), which was sutured to North America in the Triassic¹⁹. Lack of faunal control and the absence of relevant palaeomagnetic data¹⁰ preclude conclusive statements regarding the site of origin of these rocks. Thus, less direct petrological, sedimentological and structural arguments must be constructed^{1,7}.

The presence of marine phosphorites within the rock record of a particular region may have palaeogeographical implications if some analogy can be drawn between ancient and recent deposits. It is evident that many marine environments favourable to phosphate deposition in ancient oceans do not exist at present^{13,16}. However, while the oceanic and biogeochemical conditions which existed during deposition of many ancient phosphorites remain largely unexplained, several recurrent observations lead to the following conclusions:

- (1) When corrected for palaeomagnetic position, most recent and ancient phosphorites are found to have formed along the margins of continents in areas of oceanic upwelling at latitudes lower than 40° – 50° (refs 20, 21).
- (2) Phosphorites which form along the west coasts of continents ('west-coast type' deposits) in areas of trade wind belt

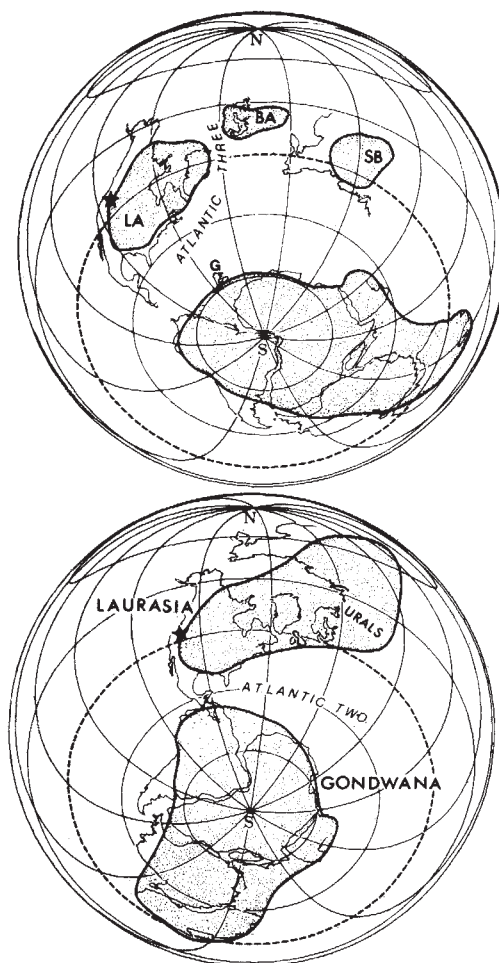


Fig. 4 Palaeomagnetic plate reconstructions for a, late Ordovician and b, late Devonian from ref. 34. ★, Present location of Sierran phosphorites with respect to North American Cordillera. Palaeoequators are shown as dashed lines.

upwelling²²⁻²⁴ and those formed along continental margins bordering narrow, east-west seaways ('Tethyan-type' deposits)²⁵ in areas of equatorial upwelling²⁶ are typically associated with black shale and chert. Phosphate within west coast- or Tethyan-type deposits is generally nodular to pelletal in form.

(3) 'Florida-type'²⁷ or 'Chatham Rise-type'²⁸ phosphate deposits, which form along the east-facing margins of continents, are not generally associated with chert and black shale but typically occur within sandstone, limestone and dolomite^{27,29,30}. Phosphate within Florida-type deposits is generally found in the form of pellets, pavements, sands and cobbles which show abundant evidence of enrichment by weathering, redeposition and replacement of limestone³⁰⁻³².

In form and lithic association phosphate deposits within Palaeozoic strata of the northern Sierra Nevada closely resemble west coast- or Tethyan-type deposits. Further, continentally derived sandstones within the SFC¹ imply relative proximity of the Sierran block to a continental mass during the early Palaeozoic. It is thus postulated that the Sierran phosphorites were deposited along the west-facing coast of a relatively large continental mass or, possibly, in a narrow east-west seaway between two continents during the early to middle Palaeozoic. By analogy with the world-wide distribution of modern and ancient phosphorites^{20,21}, the site of deposition would have been located at a latitude lower than 40°–50°. Water depths may have been in the range of 100–500 m (ref. 33).

Figure 4a, b shows the world-wide plate configuration as determined by Morel and Irving³⁴ for the general time period of phosphate deposition in the northern Sierra Nevada, approximately late Ordovician to late Devonian. The maps show that the west coast of North America was one of the few continental margins situated favourably throughout the early to middle Palaeozoic for formation of west coast-type phosphorites as most of the Gondwana continents were clustered about the South Pole. Other continental margins in positions favourable for west coast-type phosphate formation during the entire time period were the present north-west coast of Australia and northern India. The Palaeozoic stratigraphy and structural development of north-west Australia³⁵ and northern India³⁶ bear little resemblance to that of the northern Sierra Nevada^{4,7,8}. Thus, neither of these sites is believed to represent a reasonable source terrane for a Sierran microplate. The maps also show that no narrow, equatorial seaways existed during the early to middle Palaeozoic, an observation consistent with the plate reconstructions of Scotese and others³⁷.

The present position of Sierran phosphorites near the western edge of the Cordillera is consistent with the palaeogeography of the North American continent during the early to middle Palaeozoic. If the northern Sierra Nevada was in its approximate position relative to North America during the early to middle Palaeozoic then the phosphorites would have lain at approximately lat. 5°–10°N during this time period. Schweickert and Snyder³⁸ concluded that an ocean as wide as the modern Atlantic may have existed off western North America by the Middle Ordovician assuming spreading rates comparable with those along the Mid-Atlantic Ridge and an upper Proterozoic (<850 Myr) rifting event³⁹. The west-facing Palaeozoic margin of North America, at low latitudes and opposite an ocean sufficiently wide to develop circulatory gyres, was an ideal location to develop phosphatic sediments. It is thus not surprising to find phosphorites within Sierran strata if indeed they are indigenous to North America.

The hypothesis that Palaeozoic sediments of the northern Sierra Nevada were deposited on or near the Cordilleran margin of North America during the early to middle Palaeozoic has important tectonic implications. Specifically, this hypothesis is inconsistent with models which suggest that the northern Sierra Nevada represents part of a wholly exotic island arc terrane which was unrelated to North America until accretion during the Permo-Triassic Sonoma Orogeny¹⁹. Considerable translation of the northern Sierra Nevada along the continental margin¹⁹ cannot be ruled out, however, as the entire west coast of

North America could have received phosphate deposition (Fig. 4a, b). Though not conclusive, the present position of the northern Sierra Nevada within an area favourable for phosphate deposition during the early to middle Palaeozoic suggests a less complicated model wherein the Sierran phosphorites were deposited along the western margin of North America and have since experienced only limited translations. This model is consistent with the observations of other authors who have noted broad tectonic and stratigraphical similarities between Palaeozoic rocks of the Great Basin and the Sierra Nevada^{1,4,7,11,38,40-43}.

I thank G. Bond for discussions, M. Bell, P. Fenn and E. Goldish for analytical work, Union Oil Company for assistance in manuscript preparation, and W. Burnett, G. A. Crawford and E. Irving for review. Work was supported by NSF grant EAR 76-14270 awarded to E. Moores and G. Bond, University of California at Davis.

Received 2 December 1981; accepted 19 March 1982.

1. Bond, G. C. & Devay, J. J. *Geol.* **88**, 285–308 (1980).
2. Clark, L. D. *Prof. Pap. U.S. geol. Surv.* **923**, 26 (1976).
3. Schweickert, R. A. *Abstr. Prog. geol. Soc. Am.* **6**, 251 (1974).
4. D'Allura, J. A., Moores, E. M. & Robinson, L. *Pacif. Coast Paleogeogr. Symp. Soc. Econ. Paleontol. Miner.* **1**, 421–440 (1977).
5. Durrell, C. & D'Allura, J. A. *Bull. geol. Soc. Am.* **88**, 844–852 (1977).
6. Schweickert, R. A. & Girty, G. H. *Abstr. Prog. geol. Soc. Am.* **13**, 105 (1981).
7. Varga, R. J. & Moores, E. M. *Geology* **9**, 512–518 (1981).
8. Girty, G. H. & Schweickert, R. A. *Abstr. Prog. geol. Soc. Am.* **11**, 79 (1977).
9. Varga, R. J. *Abstr. Prog. geol. Soc. Am.* **11**, 133 (1977).
10. Hannah, J. L. & Verosub, K. L. *Geology* **8**, 520–524 (1980).
11. Burchfiel, B. C. & Davis, G. A. *Am. J. Sci.* **275A**, 363–396 (1975).
12. Devay, J. & Stanley, E. *Abstr. Prog. geol. Soc. Am.* **11**, 412 (1979).
13. Gulbrandsen, R. A. *Prof. Pap. U.S. geol. Surv.* **700-B**, 9–13 (1970).
14. Bendor, Y. K. *Spec. Publ. Soc. Econ. Paleont. Miner.* **29**, 3–18 (1980).
15. McArthur, J. M. *Spec. Publ. Soc. Econ. Paleont. Miner.* **29**, 53–60 (1980).
16. Bendor, Y. K. in *Rep. Marine Phosphatic Sed. Workshop*, 29 (1979).
17. Dickinson, W. R. *Pacif. Coast. Paleogeogr. Symp. Soc. Econ. Paleont. Miner.* **1**, 137–155 (1977).
18. Churkin, M. Jr. *Spec. Publ. Soc. Econ. Paleont. Miner.* **19**, 174–192 (1974).
19. Speed, R. C. *Geology* **87**, 279–292 (1979).
20. Sheldon, R. P. *Prof. Pap. U.S. GEOL. Surv.* **501C**, 106–113 (1964).
21. Cook, P. J. & McEhinny, M. W. *Econ. Geol.* **74**, 315–330 (1979).
22. McKelvey, V. E., Swanson, R. W. & Sheldon, R. P. *19th Int. Geol. Congr.* **11**, 45–65 (1953).
23. Grower, H. D. & Madsen, B. M. *Prof. Pap. U.S. geol. Surv.* **501D**, 79–85 (1964).
24. Cheney, T. M., McClellan, G. H. & Montgomery, E. S. *Econ. Geol.* **74**, 232–259 (1979).
25. Sheldon, R. P. *Econ. Geol.* **59**, 1159–1175 (1964).
26. Sheldon, R. P. *Spec. Publ. Soc. Econ. Paleont. Miner.* **29**, 239–247 (1980).
27. Cathart, J. B. *U.N. Res. Dev. Ser.*, **32**, 403–405 (1969).
28. D'Anglejan, B. F. in *Rep. Marine Phosphatic Sed. Workshop*, 17 (1979).
29. McKelvey, V. E. *Bull. U.S. geol. Surv.* **1252D**, 1–22 (1967).
30. Riggs, S. R. *Econ. Geol.* **74**, 195–220 (1979).
31. Mannheim, F. T., Pratt, R. M. & McFarlin, P. F., *Spec. Publ. Soc. Econ. Paleont. Miner.* **29**, 117–137 (1980).
32. Cullen, D. J. *Spec. Publ. Soc. Econ. Paleont. Miner.* **29**, 139–148 (1980).
33. Burnett, W. C. in *Proc. Fertilizer Raw Materials Workshop*, 120–144 (1979).
34. Morel, P. & Irving, E. J. *Geol.* **86**, 535–561 (1978).
35. Veevers, J. J. *J. geol. Soc. Aust.* **14**, 253–271 (1967).
36. Gansser, A. *Geology of the Himalayas*, 289 (Interscience, New York, 1964).
37. Scotese, C., Bambach, R. K., Barton, C., Van der Voo, R. & Ziegler, A. *J. Geology* **87**, 217–277 (1979).
38. Schweickert, R. A., & Snyder, W. S. in *The Geotectonic Development of California*, 182–202 (Prentice-Hall, New Jersey, 1981).
39. Stewart, J. H. *Bull. geol. Soc. Am.* **83**, 1345–1360 (1972).
40. Moores, E. M. *Nature* **228**, 837–842 (1970).
41. Schweickert, R. A. *Nature* **260**, 586–591 (1976).
42. Schweickert, R. A. *Abstr. Prog. geol. Soc. Am.* **9**, 497 (1977).
43. Moore, J. N. & Foster, C. T. *Bull. geol. Soc. Am.* **91**, 37–43 (1980).
44. Burchfiel, B. C. & Davis, G. A. in *The Geotectonic Development of California*, 50–70 (Prentice-Hall, New Jersey, 1981).

Temperature control of oxygen-isotope fractionation of cultured planktonic foraminifera

Jonathan Erez

The Hebrew University of Jerusalem, The H. Steinitz Marine Biology Laboratory, Eilat, PO Box 469, Israel

Boaz Luz

The Hebrew University of Jerusalem, Department of Geology, Jerusalem, Israel

The palaeotemperature equation of Epstein *et al.*¹ relates the oxygen-isotope composition of mollusc carbonate shells to the temperature in which it has been deposited. This equation has been applied extensively to derive palaeotemperatures and

other palaeoenvironmental parameters for the Pleistocene epoch, based on the isotopic composition of planktonic foraminifera in deep sea cores²⁻¹¹. Recently, doubt has been expressed as to whether planktonic foraminifera fractionate oxygen isotopes according to the known palaeotemperature equation¹²⁻¹⁸, and thus the validity of many palaeoenvironmental studies has been questioned. We present here experimental data demonstrating that planktonic foraminifera of the species *G. sacculifer* deposit calcite skeletons that have an oxygen-isotope composition very close to that predicted by the well-known palaeotemperature equation¹.

The oxygen-isotope composition of planktonic foraminifera in deep sea cores is a major tool in Pleistocene research. It serves to determine palaeotemperatures, palaeosalinities, sea-level changes, continental climatic conditions and provides accurate stratigraphy for the Pleistocene epoch²⁻¹¹. These studies are based on the assumption that oxygen isotopes in planktonic foraminifera shells are fractionated according to the equation of Epstein *et al.*¹ that was determined experimentally for molluscs. Emiliani² tested this assumption by comparing the 'isotopic temperature' (the temperature calculated based on the isotopic composition of calcium carbonate), of planktonic foraminifera from superficial sediments, to the actual temperatures of the overlying seawater. He found a good fit between these temperatures and concluded that planktonic foraminifera can be used for palaeotemperature studies. This approach has been repeated on material from sediments with similar conclusions^{6,10,19,20}. However, isotopic temperatures calculated for living planktonic foraminifera that were collected by plankton net tows were almost always significantly higher than those observed in the water in which they were collected¹²⁻¹⁸. It has been proposed that these nonequilibrium isotopic compositions are caused by introduction of metabolic CO₂ into the carbonate skeletons, a phenomenon that has been reported in other carbonate depositing organisms, including benthic foraminifera^{21,22} and in part was attributed to metabolic activity of symbiotic algae²¹. Attempts to resolve the problems using foraminifera collected in sediment traps did not reach unequivocal conclusions^{18,23}, and thus palaeoclimatic determinations based on oxygen-isotope compositions of planktonic foraminifera became uncertain.

The ability to culture planktonic foraminifera in the laboratory²⁴ opened the possibility of resolving this problem experimentally. A detailed account of the entire experiment will be reported elsewhere²⁵. Briefly, young (roughly 10-day old) individuals of the planktonic foraminifer *G. sacculifer* were collected from surface water in the Gulf of Eilat using plankton nets. Each individual was cultured in an 100-ml Erlenmeyer flask with 80 ml of filtered seawater and was fed daily by one newly hatched brine shrimp. The Erlenmeyer flasks were kept in temperature controlled baths where the deviations from the desired temperature did not exceed 0.1 °C (based on three daily accurate thermometer readings, and a continuous monitoring of temperature on strip chart recorder). Salinities were monitored every 2 days to ensure that there was no water evaporation. The growth of the foraminifera was monitored daily using an inverted microscope. All individuals considered in this experiment reached maturity and went through reproduction by producing gametes. The average initial weight of these foraminifera was ~5 µg per individual and the final weight was ~45 µg per individual. Thus ~90% of the carbonate shells was deposited in controlled conditions. A correction for the original (10% weight) isotopic composition was made using control groups for which the isotopic composition was determined separately. Each experiment lasted roughly 2 weeks; the range of experimental temperatures was between 14 and 30 °C, and in each run there were 2 or 3 different temperatures. Because only individuals that went through gametogenic reproduction were used for the final isotopic analysis, there was no need to clean the protoplasm from the shell. During the process of gametogenesis²⁴, the entire organic matter has been extruded naturally from the shell and the skeleton that is left behind is

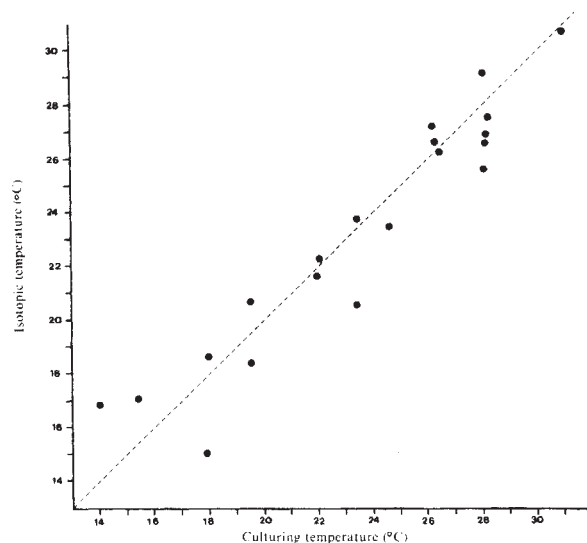


Fig. 1 The isotopic temperature calculated from the oxygen-isotope composition of the shell carbonate of the foraminifer *G. sacculifer* according to the modified equation of Epstein *et al.*¹ plotted against the actual temperature in which the foraminifera were cultured. The dashed line represents the ideal fit between the two temperatures having a slope of 1.0 and an intercept of 0. The best linear fit to the data is a line having slope and intercept of 1.01 and -0.07, respectively, having a correlation coefficient of $r = 0.95$. These experimental data demonstrate that *G. sacculifer* deposits its skeleton close to oxygen isotopic equilibrium, and that the common palaeotemperature equation for molluscs is applicable for this planktonic foraminifer.

white, organic matter free and very similar to material from sediments. Yet before analysis the individuals were crushed in ethanol, dried, vacuum roasted for 30 min at 450 °C. The analysis was carried out according to the method of Shackleton²⁶ using a VG602 Micromass spectrometer. The water isotopic composition was determined according to the procedure of Epstein and Mayeda²⁷.

The isotopic temperatures were calculated according to the equation of Epstein *et al.*¹ as modified by Craig²⁸

$$pt = 16.9 - 4.2(\delta^{18}O_c - \delta^{18}O_w) + 0.13(\delta^{18}O_c - \delta^{18}O_w)^2$$

where pt is the isotopic temperature and $\delta^{18}O_c$ and $\delta^{18}O_w$ are the isotopic composition of the carbonate and seawater respectively.

The isotopic temperature and the actual culture temperature are compared in Fig. 1. A remarkable fit exists between the isotopic and the actual temperatures. The line of best fit using linear regression has the following parameters:

$$t = -0.07 + 1.01pt$$

where t and pt are the actual and the isotopic temperatures, respectively. The correlation coefficient (r) for this linear regression equals 0.95. Ideally, the intercept and the slope should have been 0.0 and 1.0, respectively. The values that we obtained (-0.07 and 1.01) are almost identical to those expected, and would cause only a slight deviation of <1% in palaeotemperature determinations in the range of 14–30 °C.

We therefore conclude that the planktonic foraminifer *G. sacculifer* fractionate oxygen isotopes with changing temperatures according to the modified equation of Epstein *et al.*¹ that was originally developed for molluscs. Accurate palaeotemperature determinations that are based on *G. sacculifer* isotopic composition should still take into consideration the isotopic composition of seawater and the vertical migration of this species during its ontogeny^{12,23}. The nonequilibrium values that have been observed in plankton samples deserve further investigation, but it is not unlikely that they were caused by vacuum roasting of the samples before analysis, which was intended to remove organic contaminants.

Our experiment provides more data points to the palaeotemperature equation at the temperature range where it was lacking (14–30 °C) and substantiates the results of earlier palaeoenvironmental studies based on the oxygen-isotope composition of planktonic foraminifera.

We thank A. Szin and A. Shemesh for technical assistance.

Received 14 December 1981; accepted 15 March 1982.

- Epstein, S., Buchsbaum, R., Lowenstam, H. A. & Urey, H. C. *Bull. geol. Soc. Am.* **64**, 1315–1326 (1953).
- Emiliani, C. *Am. J. Sci.* **252**, 149–158 (1954); *J. Geol.* **63**, 538–578 (1955).
- Emiliani, C. *J. Geol.* **74**, 102–126 (1966).
- Shackleton, N. J. & Opdyke, N. D. *Quat. Res.* **3**, 39–55 (1973); *Geol. Soc. Am. Mem.* **145**, 449–464 (1976).
- Imbrie, J., Van Donk, J. & Kipp, N. G. *Quat. Res.* **3**, 10–38 (1973).
- Shackleton, N. J. & Vincent, E. *Mar. Micropaleont.* **3**, (1978).
- Deuser, W. G., Ross, E. H. & Waterman, L. S. *Science* **191**, 1168–1170 (1976).
- Vergnaud-Grazzini, C., Ryan, W. B. F. & Cita, M. B. *Mar. Micropaleont.* **2**, 353–370 (1977).
- Williams, D. F., Thunell, R. C. & Kennett, J. P. *Science* **201**, 252–254 (1978).
- Berger, W. H., Killingley, J. S. & Vincent, E. *Oceanol. Acta* **1**, 203–216 (1978).
- Reiss, Z. *et al. Quat. Res.* **14**, 294–308 (1980).
- Van Donk, J. thesis, Columbia Univ., New York (1970).
- Van Donk, J. *Oceanic Micropaleontology* (ed. Ramsay, A. T. S.), 1345–1370 (Academic, New York, 1977).
- Shackleton, N. J., Wiseman, J. D. & Buckley, H. A. *Nature* **242**, 177–179 (1973).
- Kahn, M. I. *Oceanol. Acta* **2**, 195–208 (1979).
- Kahn, M. I. & Williams, D. F. *Paleogeogr. Paleoclimatol. Paleocol.* **33**, 47–69 (1981).
- Duplessy, J. C., Be, A. W. H. & Blank, P. L. *Paleogeogr. Paleoclimatol. Paleocol.* **33**, 9–46 (1981).
- Deuser, W. G., Ross, E. H., Hemleben, C. & Spindler, M. *Paleogeogr. Paleoclimatol. Paleocol.* **33**, 103–127 (1981).
- Lidz, B., Kehm, A. & Miller, H. *Nature* **217**, 245–247 (1968).
- Savin, S. M. & Douglas, R. G. *Bull. geol. Soc. Am.* **84**, 2327–2342 (1973).
- Erez, J. *Nature* **273**, 199–202 (1978).
- Vinot-Bertuile, A. C. & Duplessy, J. C. *Earth planet Sci. Lett.* **18**, 247–252 (1973).
- Erez, J. & Honjo, S. *Paleogeogr. Paleoclimatol. Paleocol.* **33**, 129–156 (1981).
- Be, A. W. H. *et al. Micropaleontology* **23**, 155–179 (1977).
- Erez, J. & Luz, B., *Geochim. cosmochim. Acta* (submitted).
- Shackleton, N. J. *Colloq. int. cent. natn. Rech. scient.* No. 219, 203–209 (1973).
- Epstein, S. & Mayeda, T. *Geochim. cosmochim. Acta* **4**, 213–224 (1953).
- Craig, H. *Proc. Spoleto Conf. on Stable Isotopes in Oceanographic Studies and Paleotemperatures* Vol. 3 (ed. Tongiorgi, E.) (1965).

IQ in Japan and the United States shows a growing disparity

Richard Lynn

Department of Psychology, The New University of Ulster, Coleraine, Co. Londonderry BT52 1SA, UK

Evidence from 27 samples indicates that the mean IQ in Japan is higher than in the United States by around one-third to two-thirds of a standard deviation. Analysis of results from the standardization in Japan in 1975 of the new revised version of the American Wechsler Intelligence Scale for Children shows that the Japanese–American disparity in mean IQ has increased during the twentieth century. Among the younger generation the mean Japanese IQ is approximately 111.

Several studies have indicated that the mean IQ is higher in Japan than in the United States and other advanced western nations^{1,2}. The Japanese–American disparity has varied in different samples from 2 to 15 IQ points. The existing evidence

suggests that the disparity is generally smallest among samples born in the earlier decades of the century, and greatest among young samples born more recently. Thus the mean IQ in Japan may have risen relative to that in the United States, producing an increasing disparity between the two populations.

Evidence to determine whether this is the case can be derived from the recent standardization in Japan of the revised version of the Wechsler Intelligence Scale for Children³. This test was constructed and standardized in the United States in the early 1970s and was standardized in Japan in 1975. Five of the performance subtests were retained unaltered in the Japanese standardization—Picture Arrangement, Object Assembly, Coding, Mazes and Block Design (in Block Design the last item in the Japanese version has been made slightly more difficult, thus making the test harder for the Japanese). The Digit Span subtest was also unchanged in the Japanese standardization, as in the previous standardizations of the Wechsler tests.

In Japan, the test was standardized on a representative sample of the child population consisting of 100 subjects at each age level from 6–16 yr. The standardization procedure in Japan followed that in the United States in matching the sample for socio-economic status to the total population. The mean raw scores of the Japanese samples at each age can be read off from the Japanese manual, and then converted to American scaled scores and the corresponding IQs ascertained from the American manual. Table 1 gives the mean Japanese scaled scores on the five performance subtests and on Digit Span, and the mean performance IQ derived from the five subtests. The overall mean performance IQ of this Japanese sample is 111 (relative to a mean American IQ of 100). 1,100 Japanese and 2,200 Americans were tested and the difference in mean IQ is clearly statistically significant. For individual tests for each year, a difference of one scaled point is statistically significant at the 0.01 probability level. The result evidently confirms previous estimations indicating that the Japanese mean IQ is higher than that in the United States.

Examination of the scores obtained by the Japanese on the six subtests given in Table 1 shows that the Japanese superiority is most pronounced in the tests of block design, mazes, picture arrangement and object assembly. The Japanese are less good on the digit span and coding tests. These results are broadly similar to those found in the Japanese standardization of the WISC and the WAIS. The consistency of this pattern of results suggests that the Japanese may be particularly strong on spatial ability.

Combining the results from the WISC-R with those of previous studies we have IQ data for 27 Japanese cohorts spanning almost seven decades. The series starts with the results for the WAIS for which the Japanese standardization sample was born around 1910 and concludes with the 6-yr-olds in the WISC-R born in 1969. Table 2 gives the mean IQs of the 27 cohorts in relation to their year of birth. It is clear that the Japanese cohorts born earlier in the century (1910–45) have a mean IQ of 102–105, whereas those born from 1946 to 1969 have a mean of 108–115. This suggests that the mean Japanese IQ has been rising relative to the American during the twentieth century. The secular trend is shown in Fig. 1, in which mean

Table 1 Japanese mean WISC-R scaled scores and IQs based on US norms

Age	Digit span	Picture arrangement	Block design	Object assembly	Coding	Mazes	Performance IQ
6	10.5	11.8	13.9	11.6	9.0	13.2	112
7	10.5	11.7	14.2	11.7	7.7	13.0	111
8	10.7	10.8	14.4	11.5	9.4	12.7	112
9	9.8	12.0	14.9	11.0	9.2	12.0	112
10	10.7	11.3	14.7	11.7	9.2	11.7	111
11	10.7	11.0	14.9	11.7	9.3	11.5	111
12	10.7	11.0	14.3	11.7	9.7	11.8	112
13	10.0	11.0	14.1	11.7	9.9	11.3	111
14	10.8	11.0	14.2	11.0	9.8	12.0	111
15	10.5	11.0	14.2	11.0	9.8	11.0	109
16	11.5	11.0	13.5	11.0	9.3	10.0	106

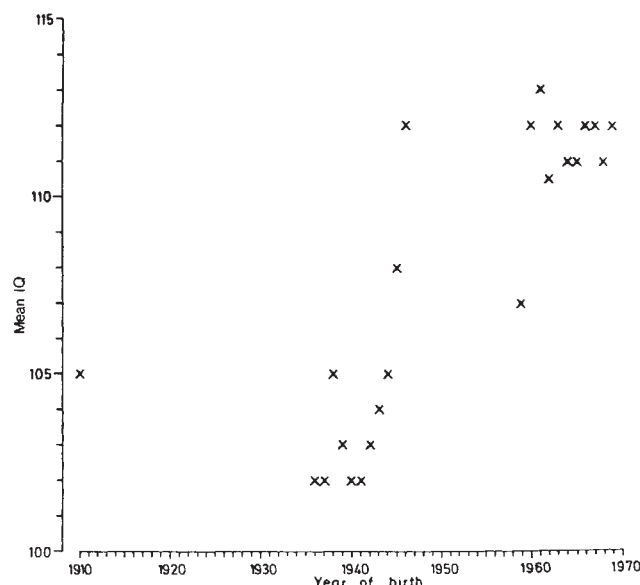


Fig. 1 Mean IQs of 23 Japanese cohorts showing a secular increase relative to a mean IQ of 100 in the US.

IQs for the same cohort are averaged to give a single mean (this applies to the four cohorts born during 1959–62).

The significance of the rising trend of the mean IQ in Japan can be tested from the correlation between mean IQ and year of birth. The product moment correlation is $+0.76$ and is statistically significant at $P < 0.01$. The greatest discrepancy is in the value for 1910, which was derived from the WAIS and is based on only two of the performance subtests and digit span. It may therefore be subject to error. If this value is discounted, the cohorts born in the years 1936–46 have a mean IQ of 104.34, while those born in the years 1959–62 have a mean IQ of 111.20. The value of χ^2 for this difference is 10.83 and is statistically significant at $P < 0.01$ (Yates' correction is applied).

Thus over the course of a generation the mean IQ in Japan has risen by ~ 7 IQ points. It seems doubtful whether a rise of this magnitude could be accounted for by a change in the genetic structure of the population. Instead, the explanation probably lies largely in environmental improvements. Comparison of the WISC and the WISC-R shows that the increase in IQ was present among 6-yr-olds. This indicates that it cannot be explained in terms of improvements in education and must be attributed to effects taking place before the age of six. Improvements in health and nutrition may be involved as it has been shown that the birth weight of Japanese babies has increased over the middle decades of the century⁴.

Other advanced western nations, such as Britain, France, Belgium, Germany, Australia and New Zealand, all have mean

IQs approximately the same as that in the United States⁵. At 111, the mean IQ in Japan is the highest recorded for a national population by a considerable order of magnitude. The effects of the high Japanese IQ of 111 can be illustrated as follows. Whereas Americans and Europeans have $\sim 2\%$ of their populations with IQs over 130, the Japanese have $\sim 10\%$ at this level. Among the population as a whole, 77% of Japanese have a higher IQ than the average American or European. Since intelligence is a determinant of economic success, as it is of success in many other fields, the Japanese IQ advantage may have been a significant factor in Japan's outstandingly high rate of economic growth in the post-World War II period.

Received 30 November 1981; accepted 26 March 1982.

1. Lynn, R. *Bull. Br. psychol. Soc.* **30**, 69 (1977).
2. Lynn, R. & Dziobon, J. *Person. Indiv. Diff.* **1**, 95 (1980).
3. Kodama, H., Shinagawa, F. & Motegi, M. *Wechsler Intelligence Scale for Children—Revised* (Nihon Bunka Kagakusha, 1978).
4. Gruenewald, P., Funakawa, H., Mitani, S., Nishimura, T. & Takeuchi, S. *Lancet*, 1026–1028 (1967).
5. Lynn, R. in *Human Variation* (eds Osborne, R. T., Noble, G. E. & Weyl, N.) (Academic, New York, 1978).

Light experience and asymmetry of brain function in chickens

Lesley J. Rogers

Pharmacology Department, Monash University, Wellington Road, Clayton, 3168 Victoria, Australia

Asymmetry of brain function has been demonstrated for several non-human species¹, most clearly in the avian brain^{2,3}. In the chicken, a single treatment of the left forebrain hemisphere during the first week of life after hatching with either cycloheximide or the putative neurotransmitter, glutamate, subsequently causes retarded visual and auditory learning, and elevated attack and copulation responses^{3–5}. Similar treatment of the right hemisphere is without effect. The cellular mechanisms by which these pharmacological agents may alter brain development and so reveal lateralization have already been discussed in detail^{6,7}. It is commonly assumed that lateralization of brain function in humans and other species is inherited either genetically^{1,8} or cytoplasmically⁹. However, Rogers and Anson have suggested previously³ that light experience may have an important role in establishing lateralization in the chicken forebrain, because after day 17 of incubation the embryo is oriented in the egg such that the left eye is occluded by the chicken's wing and body while the right eye is next to the air sac and exposed to light input (refs 10, 11 and J. V. Zappia and L.J.R., in preparation). Because the optic nerves decussate completely and most of the information reaching each tectum is processed by its ipsilateral forebrain hemisphere, it is possible that light entering the right eye stimulates developmental processes in the left hemisphere in advance of the right³. If so, chickens hatched from eggs incubated in darkness should fail to show functional asymmetry of the forebrain. I now report that this is indeed the case for asymmetrical control of attack and copulation.

Because all previous experiments demonstrating asymmetrical control of attack and copulation have used only male chickens hatched from eggs exposed to light during incubation^{4,5}, it was first necessary to ascertain that the same phenomenon occurred in females also hatched from eggs exposed to light, since chicks hatched in my laboratory were not sexed. As found for males, glutamate (500 nmol in 5 μ l) treatment of the left hemisphere of female chicks on day 2 after hatching caused elevated attack and copulation (see ref. 4 for details of method), whereas those treated similarly in the right hemisphere behaved the same as untreated controls (Fig. 1).

Table 2 Mean IQs of 23 Japanese cohorts and their year of birth

Year of birth	Mean IQ	Test	Year of birth	Mean IQ	Test
1910	105	WAIS	1959	108	WPPSI
1936	102	WISC		106	WISC-R
1937	102	WISC	1960	115	WPPSI
1938	105	WISC		109	WISC-R
1939	103	WISC	1961	115	WPPSI
1940	102	WISC		111	WISC-R
1941	102	WISC	1962	110	Kyoto
1942	103	WISC		111	WISC-R
1943	104	WISC	1963	112	WISC-R
1944	105	WISC	1964	111	WISC-R
1945	108	WISC	1965	111	WISC-R
1946	112	WISC	1966	112	WISC-R
			1967	112	WISC-R
			1968	111	WISC-R
			1969	112	WISC-R

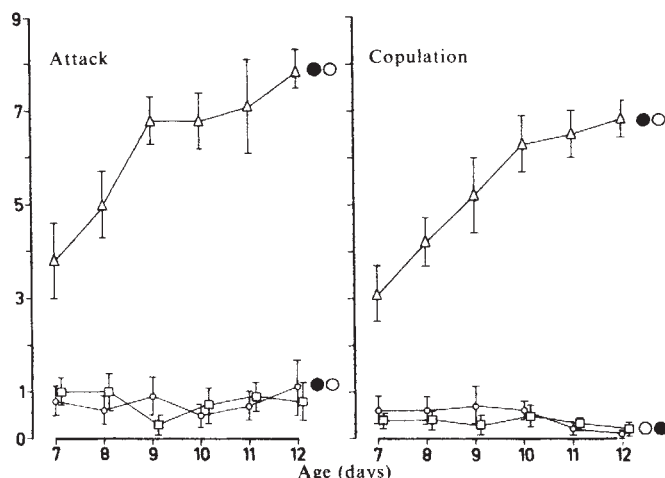


Fig. 1 Female chicks, supplied by a hatchery in which the eggs are exposed to light during incubation, have been tested for attack and copulation from days 7 to 12 of life. Daily mean scores with standard errors are presented. The behaviours are scored according to a rank ordering procedure (0–10) in standard tests using the moving hand as a stimulus⁴. Scores plotted on the y-axis rank from no response (0) to sparring with attack leaps for attack, and treading with pelvic thrusting for copulation. Δ Represents females treated with glutamate in the left hemisphere, and on every day of testing their scores are significantly ($P < 0.002$, *U*-tests) elevated above those for females treated in the right hemisphere ($\circ$), and untreated females ($\square$). The pair of circles next to each graph represents the hemispheres, the hemisphere treated with glutamate being solid black. The other hemisphere is treated with normal saline. There were 8–10 animals in each group.

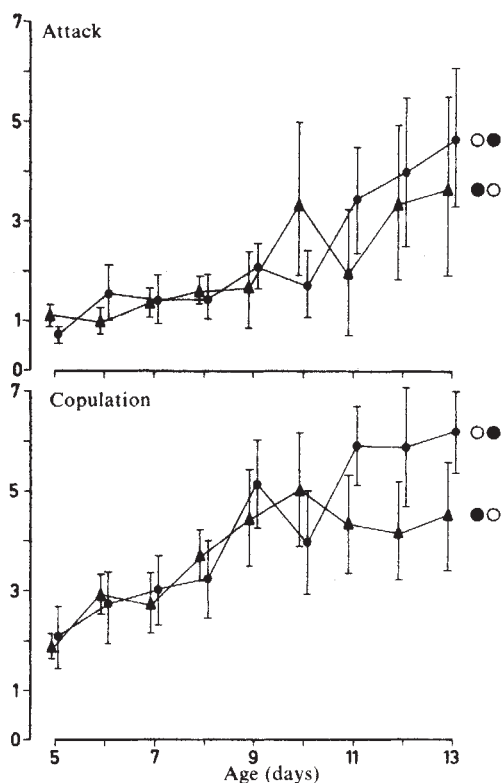


Fig. 2 Attack and copulation scores for chicks hatched from eggs incubated completely in darkness. Symbols as in Fig. 1; $\blacktriangle$, glutamate treatment in the left hemisphere ($n = 15$); $\bullet$, glutamate treatment in the right ($n = 10$).

Male and female chickens hatched from eggs incubated in darkness and treated with 500 nmol of glutamate on day 2 after hatching show no asymmetry of function for attack and copulation (Fig. 2). As the mean scores in Fig. 2 demonstrate, the chicks in both the left and right hemisphere treatment groups showed some elevation of attack and copulation scores.

In the chicken the first visually evoked tectal responses occur on day 17 of incubation. An electroretinogram of embryonic form can be detected and the first responses to light can be measured on day 18, and on day 19 the electroretinogram reaches its post-hatch form and light-evoked potentials are detected in the contralateral forebrain hemisphere^{11,12}. Thus, it is during the last 5 days of incubation that light exposure should affect brain asymmetry, and this is in fact the case. In a matched control experiment, one batch of eggs was incubated in darkness for the first 16 days of incubation and then exposed to light of intensity varying between 100 and 500 lx during the last 5 days; another batch received the reverse treatment. The former treatment causes lateralization of attack and copulation, the latter does not (see Fig. 3).

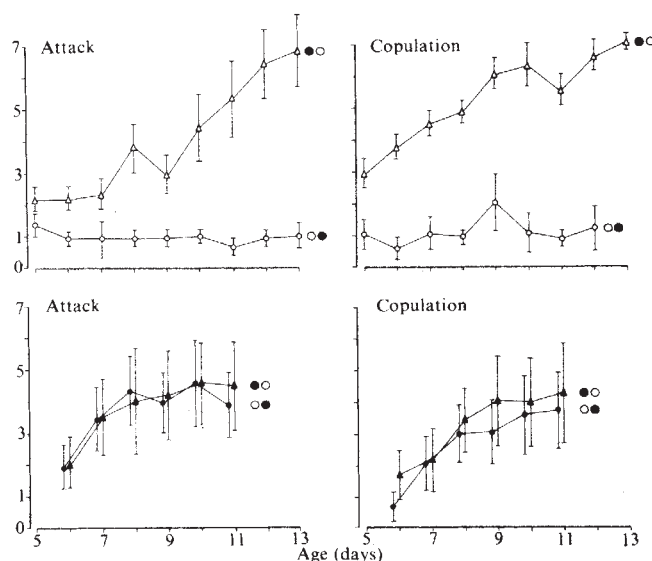


Fig. 3 The two graphs at the top present data for chicks hatched from eggs which were incubated in darkness for the first 16 days of incubation, and light from a heating bulb and overhead lamps (100–500 lx) during the last 5 days. The graphs at the bottom present data for eggs incubated in the reverse conditions ($n = 8$ –10 for each group). Δ , $\blacktriangle$, Treated with glutamate in the left hemisphere; $\circ$, $\bullet$, treated in the right hemisphere.

In fact, 2 h of light exposure on day 19 of incubation are sufficient to establish asymmetrical control of attack and copulation (see Fig. 4). A sensitive period for this effect seems to have passed by the time the chick pips the egg shell on day 20 because no asymmetry for attack and copulation was present in chickens hatched from eggs exposed to light from the time of pipping onwards (~ 24 h).

As shown by the standard errors in the figures, the data for those groups of chickens hatched from eggs which had not received light exposure during the sensitive period were more variable than those of the lateralized groups ($P \leq 0.001$ for a comparison of the attack and the copulation scores on days 11 or 12 of all the lateralized versus non-lateralized chicks treated in the left hemisphere, and $P < 0.05$ for both attack and copulation scores in a similar lateralized versus non-lateralized comparison of all those treated in the right hemisphere; Moses tests¹³, $n = 38$ –48 for each group). Examination of the individual scores revealed that approximately half the chicks hatched from eggs kept in darkness during the sensitive period and then treated in the left hemisphere had consistently high scores for attack and copulation clustered around 7.0, while the other half had consistently low scores clustered around 1.0,

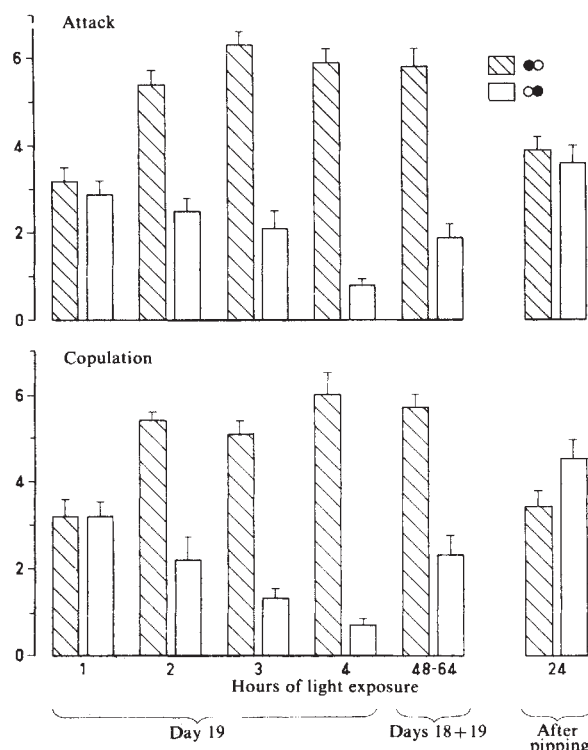


Fig. 4 Attack and copulation scores are given as overall means ($\pm$ s.e.) calculated from scores recorded daily from days 6 to 11 of life. The symbols are as in Figs 1 and 2. Apart from the hours of light exposure indicated, the eggs were incubated in darkness. There is a significant elevation ($P < 0.002$) in both scores for the left-treated groups (white columns) compared with the right-treated groups (black columns) after 2, 3, 4 and 48–64 h light exposure on day 19 of incubation, but not after 1 h on the same day or ~ 24 h from the time of pipping to hatching ($n = 8$ –12 for each group).

which suggests a bimodal distribution. This was also the case for chicks from eggs incubated similarly and treated in the right hemisphere. Further sampling will be necessary to confirm this (J. V. Zappia and L.J.R., in preparation), but it is possible that in the absence of light exposure before hatching individual chickens still have lateralized forebrain function, but that half are asymmetrical in one direction and half in the other direction, similar to pawedness in mice¹⁴. Light exposure during embryonic development may not generate lateralization in individuals, but synchronize the direction of lateralization in the population such that differences emerge between groups treated with glutamate in the left or right hemisphere.

On about day 19 of incubation, when the visual pathways between the tectum and the Wulst area of the forebrain are becoming functional, asymmetrical light input may synchronize the direction of functional asymmetry in the population, perhaps by stimulating blood flow to the left hemisphere in response to neuronal activity¹⁵ and increasing protein synthesis¹⁶ or some other metabolic process. The direction of asymmetry for attack and copulation therefore seems to be a consequence of the orientation of the embryo in the egg.

Denenberg¹⁷ has reported that asymmetry of hemispheric function in the rat is enhanced by the experience of handling in early life. In this case environmental experience influences the degree of asymmetry, differing from the chicken data in which light experience seems to affect the direction but not the degree of asymmetry. However, both examples illustrate that interaction between genotype and experience determines the presence or form of functional asymmetry in the brain.

Received 5 January; accepted 26 March 1982.

- Denenberg, V. H. *Behav. Brain Sci.* **4**, 1–21 (1981).
- Nottebohm, F. in *Lateralization in the Nervous System* (eds Harnad, S., Doty, R. W., Goldstein, L., Jaynes, J. & Kravthamer, G.) 23–44 (Academic, New York, 1977).

- Rogers, L. J. & Anson, J. M. *Pharmac. Biochem. Behav.* **10**, 679–686 (1979).
- Howard, K. J., Rogers, L. J. & Boura, A. L. *Brain Res.* **188**, 369–382 (1980).
- Rogers, L. J. *Acta XVII congr. int. orn.* **1**, 653–659 (1980).
- Hambley, J. W. & Rogers, L. J. *Neuroscience* **4**, 677–684 (1979).
- Rogers, L. J. & Hambley, J. W. *Behav. Brain Res.* **4**, 1–18 (1982).
- Levy, J. in *Hemisphere Function in the Human Brain* (eds Dimond, S. J. & Beaumont, J. G.) 121–177 (Elek, London, 1974).
- Morgan, M., in *Lateralization in the Nervous System* (eds Harnad, S., Doty, R. W., Goldstein, L., Jaynes, J. & Kravthamer, G.) 173–194 (Academic, New York, 1977).
- Oppenheim, R. W. in *Studies on the Development of Behaviour and the Nervous System* Vol. 1 (ed. Gottlieb, G.) 163–244 (Academic, New York, 1978).
- Freeman, B. M. & Vince, M. A. *Development of the Avian Embryo* (Chapman & Hall, London, 1974).
- Sedláček, J. *Adv. Psychobiol.* **1**, 129–170 (1972).
- Daniel, W. W. *Applied Non-parametric Statistics* (Houghton Mifflin, Boston, 1978).
- Collins, R. L. *J. Hered.* **60**, 117–119 (1969).
- Bondy, S. C. & Harrington, M. E. *Science* **199**, 318–319 (1978).
- Bateson, P. P. G., Horn, G. & Rose, S. P. R. *Brain Res.* **39**, 449–465 (1972).
- Denenberg, V. H., Garbanati, J., Sherman, G., Yutzy, D. A. & Kaplan, R. *Science* **201**, 1150–1151 (1978).

Is oxytocin an ovarian hormone?

D. Claire Wathes & R. W. Swann

Department of Anatomy, The Medical School, University of Bristol, Bristol BS8 1TD, UK

Oxytocin is a nonapeptide hormone produced by the hypothalamus and released from the posterior pituitary. It is cleaved from a large molecular weight precursor which is synthesized in cell bodies of hypothalamic magnocellular neurones and then packaged into membrane-bound granules¹. During axonal transport the precursor is cleaved to produce oxytocin and an oxytocin-related neurophysin. Apart from its established roles in lactation and labour, oxytocin is thought to be an important regulator of the oestrous cycle; it causes luteolysis in cattle² and immunization against oxytocin increases the length of the ovine oestrous cycle³. McCracken⁴ has proposed that oxytocin acts by stimulating prostaglandin $F_{2\alpha}$ ($PGF_{2\alpha}$) release from the uterine endometrium and that steroid hormones regulate $PGF_{2\alpha}$ secretion by altering the availability of endometrial oxytocin receptors^{4,5}. Steroids may also have a direct influence on oxytocin release from the posterior pituitary, oestradiol-17 β increasing and progesterone inhibiting oxytocin secretion in response to vaginal distension in the goat^{6,7}. However, recent radioimmunoassay (RIA) measurements of peripheral oxytocin levels during the ovine oestrous cycle^{8–10} consistently show that oxytocin and progesterone concentrations increase and decrease in synchrony during the luteal phase, reaching basal levels after luteolysis at a time when oestradiol-17 β titres are known to rise¹¹. Therefore, there may exist different controls of reflex release and basal secretion of oxytocin, and our discovery of large amounts of oxytocin within the ovine corpus luteum reported here may provide an explanation.

Peptides were extracted from ovine corpora lutea using the method developed by Walsh and Niall¹² for the isolation of porcine relaxin. The crude extract was applied to a Sephadex G-50 column. While testing the fractions from this column for relaxin-like activity in a rat uterine strip bioassay, it became apparent that some fractions were stimulatory rather than inhibitory. A similar finding has been reported previously in the cow¹³. The fractions were therefore measured by RIA for cross-reactivity with oxytocin (Fig. 1). A single peak of oxytocin-like immunoreactivity eluted between 330 and 400 ml in a position corresponding to that of synthetic oxytocin (Syn-tocinon, Sandoz). Dilution curves of material from this peak (henceforth referred to as luteal extract) showed parallelism with the oxytocin standard curve (Fig. 2). HPLC analysis of the extract revealed oxytocin immunoreactivity eluting in a position identical to that of synthetic oxytocin (Fig. 3).

Luteal extract was tested for oxytocin in two bioassays using concentrations calculated from the RIA results to correspond to the oxytocin standards. In a rat milk ejection assay, close

quantitative and qualitative agreement was achieved between the response to both luteal extract and oxytocin standard. In the rat uterine strip bioassay the estimated oxytocin-like activity approximated to one third of that measured by RIA but this discrepancy may have been due to the presence of a factor which inhibited uterine activity in the same Sephadex G-50 extract.

These data clearly indicate that the peptide extracted from the corpora lutea is authentic oxytocin. The yield of oxytocin from the ovaries of non-pregnant sheep was 2.6 µg per g of luteal tissue whereas the yield from pregnant sheep was only 34 ng per g. As the average weight of ovine corpora lutea is ~0.5 g and because there are normally 1–3 corpora lutea per animal, a non-pregnant ewe in the luteal phase of the oestrous cycle has ~2.6 µg of ovarian oxytocin. Lederis¹⁴ calculated that the posterior pituitary and hypothalamus of the sheep together contained 18.2 µg oxytocin. Thus ovarian oxytocin represents

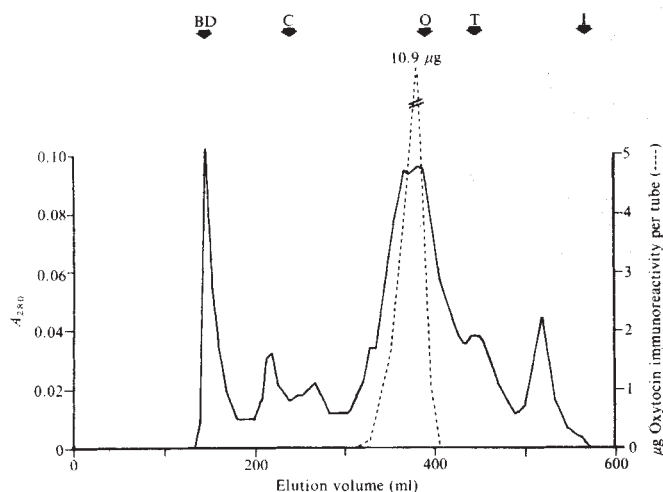


Fig. 1 Elution profile of a crude extract of 14.5 g ovine corpora lutea applied to a Sephadex G-50 (fine) column (120×2.4 cm) measured as absorbance at 280 nm. The eluant used was 0.1 M ammonium acetate pH 6.8 and the flow rate was 20.1 ml h⁻¹. Fractions (6.7 ml) were collected and lyophilized to remove ammonium acetate. The elution positions of various markers are shown: BD, blue dextran; C, cytochrome c (MW 12,300); O, oxytocin (MW 1,000); T, tyrosine (MW 181); I, ¹²⁵I. Paired fractions were assessed for oxytocin immunoreactivity by RIA. The rabbit antiserum used was raised against oxytocin coupled to thyroglobulin by the technique of Sofroniew *et al.*²⁴ and showed no significant cross-reaction with arginine-vasopressin, arginine-vasotocin or the rat neurophysins. ¹²⁵I-oxytocin, prepared by the chloramine-T method, was purified by chromatography on Sephadex G-25. The 4th International Standard for oxytocin (76/575) was used as a reference preparation.

~15% of the total hypothalamo-neurohypophysial store of hormone in the non-pregnant ewe during the luteal phase and ~0.2% during pregnancy.

Although it is possible that the presence of oxytocin in the ovary reflects its uptake from the blood, the high concentration of hormone suggests that it is synthesized there. However, proof of this hypothesis awaits the results of further experiments such as incorporation studies. If oxytocin is synthesized in a similar manner in the ovary and brain, one would also expect to find an ovarian neurophysin. We were unable to assay for ovine neurophysin but the Sephadex G-50 peak expected to contain molecules of molecular weight (MW) 10,000 (200–295 ml, Fig. 1) did contain small amounts of a component having the electrophoretic and HPLC characteristics of bovine neurophysin I.

Luteal oxytocin is probably not important for lactation in the ewe, as corpora lutea regress at parturition and a lactating ewe may not ovulate again for several months. Ovariectomized sheep can also deliver successfully, arguing against a role in parturition^{15,16}. However, oxytocin measured in the peripheral

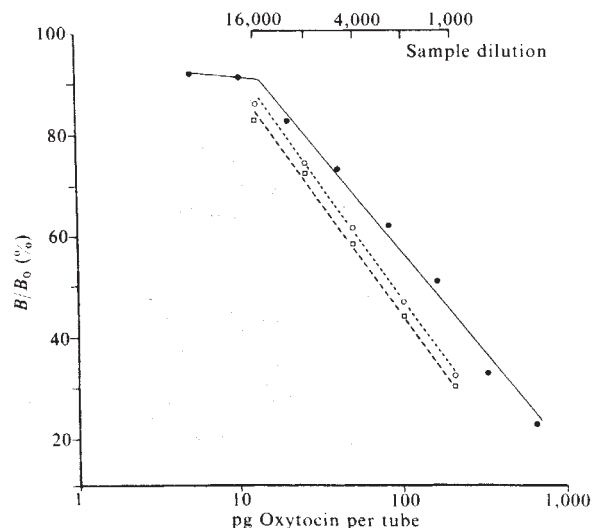


Fig. 2 Dilution curves for oxytocin standard (●) and two Sephadex-G50 fractions taken from separate extractions (○ and □). The ordinate shows bound ¹²⁵I-oxytocin (*B*) expressed as a percentage of ¹²⁵I-oxytocin bound in the absence of oxytocin (*B*₀). The slopes of the lines for the standards and samples did not differ significantly.

plasma during the oestrous cycle^{8–10} could be of luteal origin as the increases and decreases in its concentration correspond to the waxing and waning of the corpus luteum. Additional evidence in support of this view is that ovariectomy leads to a decrease in circulating oxytocin¹⁰ and that episodic release was not detected during seasonal anoestrus⁹. Circulating oxytocin concentrations also decline during pregnancy^{8,9} corresponding to the much lower yield of oxytocin obtained from the ovaries of pregnant ewes in the present study.

The ovine corpus luteum is known to contain secretory granules, the release of which reaches a maximum in mid-cycle^{17,18}. It has been suggested that these granules contain progesterone^{17,19,20}, but steroid secretion is not usually associated with granule formation and release. The present findings raise the possibility that these granules represent the ovarian

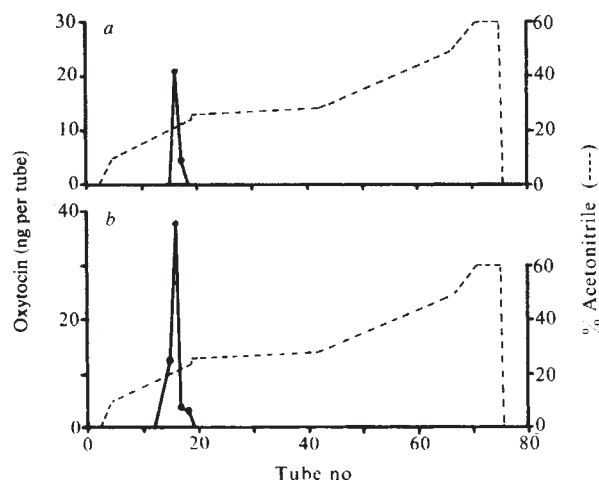


Fig. 3 Profiles obtained by adding oxytocin standard (*a*) or ovine luteal extract (*b*) to an HPLC system (Altex Model 322) and measuring aliquots of each fraction by oxytocin RIA. Chromatography was performed on a 5 mm (i.d.)×10 cm stainless steel column packed in our laboratory with Hypersil ODS 5 µm. Samples were loaded and washed on to the column in 0.2 M Na₂HPO₄ pH 2.1 (adjusted with H₃PO₄) and eluted at a flow rate of 1 ml min⁻¹ with an increasing gradient of acetonitrile (HPLC grade S, Rathburn). Fractions (1.2 ml) were collected. Vasotocin and arginine-vasopressin elute in tubes 11 and 12, respectively, using this gradient system.

oxytocin stores. Indeed another peptide hormone, relaxin, is packaged by the luteal cells into similar granules during pregnancy in the pig and rat^{21,22}.

The discovery of oxytocin in the corpus luteum raises a number of questions. If oxytocin were principally involved in the control of PGF_{2α} release at luteolysis, one would expect its concentration to peak shortly before the normal rise in PGF_{2α}, but oxytocin levels decrease at about this time⁸. McCracken⁴ has proposed that a rise in uterine oxytocin concentrations induced by oestradiol-17β is the first step in luteolysis but oestrogen levels similarly do not rise until progesterone concentrations have started to decline²³. Thus the role of oxytocin during the oestrous cycle remains undetermined.

Received 16 February; accepted 24 March 1982.

- Pickering, B. T. *Essays Biochem.* **14**, 45–81 (1978).
- Armstrong, D. T. & Hansel, W. J. *Dairy Sci.* **42**, 533–542 (1959).
- Sheldrick, E. L., Mitchell, M. D. & Flint, A. P. F. *J. Reprod. Fert.* **59**, 37–42 (1980).
- McCracken, J. A. *Adv. Prostaglandin Thromb. Res.* **8**, 1329–1344 (1980).
- Roberts, J. S., McCracken, J. A., Gavagan, J. E. & Soloff, M. S. *Endocrinology* **99**, 1107–1114 (1976).
- Roberts, J. S. *Endocrinology* **89**, 1029–1033 (1971).
- Roberts, J. S. *Endocrinology* **89**, 1137–1141 (1971).
- Webb, R., Mitchell, M. D., Falconer, J. & Robinson, J. S. *Prostaglandins* **22**, 443–453 (1981).
- Sheldrick, E. L. & Flint, A. P. F. *Prostaglandins* **22**, 631–636 (1981).
- Schams, D., Lahlou-Kassi, A. & Glatzel, P. J. *Endocr.* **92**, 9–13 (1982).
- Baird, D. T. & Scaramuzzi, R. J. *J. Endocr.* **70**, 237–245 (1976).
- Walsh, J. R. & Niall, H. D. *Endocrinology* **107**, 1258–1260 (1980).

We thank Mr R. Jones, Miss J. Creek, Miss A. Moffatt and Mr P. Rees for technical assistance; S. D. Birkett for the radioimmunoassay; Dr J. A. Wakerley for the milk ejection experiment; Dr C. M. Wathes for the statistical analysis and Professors B. T. Pickering and D. G. Porter for their helpful discussions. The work was supported in part by grants from the ARC (AG 7/103), MRC (G8011540SA, G979/479) and the Wellcome Trust (9267/1.4P).

Note added in proof:

HPLC and RIA studies using bovine corpora lutea have shown that both oxytocin and bovine neurophysin I are present in the ovary in this species too.

- Fields, M. J., Fields, P. A., Castro-Hernandez, A. & Larkin, L. H. *Endocrinology* **107**, 869–876 (1980).
- Lederis, K. *Gen. comp. Endocr.* **1**, 80–89 (1961).
- Neher, G. M. & Zarrow, M. X. *J. Endocr.* **11**, 323–330 (1954).
- Fylling, P. *Acta. endocr.* **65**, 273–283 (1970).
- Gemmel, R. T., Stacey, B. D. & Thorburn, G. D. *Biol. Reprod.* **11**, 447–462 (1974).
- Paavola, L. G. & Christensen, A. K. *Biol. Reprod.* **25**, 203–215 (1981).
- Gemmel, R. T. & Stacy, B. D. *Am. J. Anat.* **155**, 1–14 (1979).
- Quirk, S. J., Wilcox, D. L., Parry, D. M. & Thorburn, G. D. *Biol. Reprod.* **20**, 1133–1145 (1979).
- Kendall, J. Z., Plopper, C. G. & Bryant-Greenwood, G. D. *Biol. Reprod.* **18**, 94–98 (1978).
- Anderson, M. L. & Long, J. A. *Biol. Reprod.* **18**, 110–117 (1978).
- Baird, D. T., Land, R. B., Scaramuzzi, R. J. & Wheeler, A. G. *J. Endocr.* **69**, 275–286 (1976).
- Sofroniew, M. V., Madler, M., Müller, O. A. & Scriba, P. C. *Z. analyt. Chem.* **290**, 163 (1978).

Noradrenergic modulation of dendrodendritic inhibition in the olfactory bulb

C. E. Jahr* & R. A. Nicoll

Departments of Pharmacology and Physiology,
University of California, San Francisco, California 94143, USA

Noradrenaline containing fibres in the central nervous system (CNS) have been characterized in detail by histochemical techniques^{1–3}. A particularly striking feature of this neurotransmitter system is its diffuse pattern of innervation. Although noradrenaline generally inhibits spontaneous activity of neurones^{4–7}, in a number of systems^{5,8–10} it can actually enhance orthodromic excitatory responses. Enhancement of excitatory transmission through sensory relay neurones is consistent with the view that noradrenergic systems are intimately involved in such behaviours as arousal and attention¹¹. However, the basis for this enhancement is unclear. Two proposed hypotheses are a facilitation of excitatory transmitter action⁵ and a disinhibitory action in which noradrenaline inhibits inhibitory interneurons⁸. The olfactory bulb is ideally suited for studying the cellular action of noradrenaline because the neuronal circuitry is relatively simple. Using intracellular recording techniques we have now found that noradrenaline attenuates the inhibitory feedback onto relay neurones, thus facilitating their firing, while glutamate and aspartate augment the inhibitory feedback.

The olfactory bulb has two major cell types: mitral cells, which relay signals to the cortex and granule cells, which are inhibitory interneurons. The interaction between these two types of cells involves a reciprocal dendrodendritic inhibitory pathway^{12–18}. Briefly, impulses in mitral cell dendrites release excitatory transmitter (probably aspartate¹⁹) onto the dendrites of granule cells which, in turn, release the inhibitory transmitter, γ-aminobutyric acid (GABA) back onto mitral cells. In addition, the olfactory bulb receives a massive input of centrifugal fibres²⁰, some of which contain noradrenaline²¹. As some of the centrifugal fibre synapses contain spherical vesicles similar to those found in mitral cells, the acidic amino acids glutamate and aspartate may also be neurotransmitters in centrifugal

fibres. The observed close approximation of the centrifugal synapses to the reciprocal synapses suggests that they may act to change the gain of the reciprocal inhibitory pathway.

Standard intracellular recording techniques were used in the present study and are described in detail elsewhere^{16,17}. Briefly, turtles (*Pseudemys scripta elegans*) were decapitated and the olfactory bulbs removed and hemisected. The superfusate was a modified Ringer solution buffered with Tris(hydroxymethyl)aminomethane to pH 7.4 and bubbled with 100% O₂. Experiments were performed at 19°C. The olfactory nerves and mitral cell axons were stimulated with either concentric bipolar or bipolar needle electrodes. Drugs were applied in the superfusate or by iontophoresis. The placement of the tip of

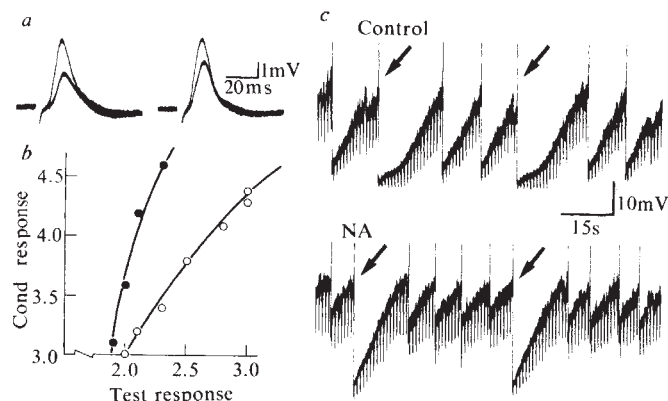


Fig. 1 Effect of noradrenaline (NA) on synaptic inhibition of mitral cells. *a*, Extracellular field potentials recorded in the granule layer to mitral cell axon stimulation. The superimposed records on the left show the conditioning response and the test response elicited 8 s later. The records on the right were obtained 20 min after switching to a Ringer containing 5 μM noradrenaline. Positivity is up in this and the following figures. *b*, Graph of responses partially illustrated in *a*. The amplitude of the conditioning response (ordinate) is plotted against the amplitude of the test response (abscissa) obtained with a series of stimulus strengths. ●, Control responses; ○, responses obtained in the presence of noradrenaline. *c*, Top trace: continuous intracellular record of orthodromic (arrows) and spontaneous i.p.s.p.s in control superfusate. Lower trace: the same, 12 min after the addition of 50 μM noradrenaline. The downward deflections are responses of the membrane to constant current hyperpolarizing current pulses. The threshold for this cell was –50 mV.

* Present address: Department of Neurobiology, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA.

the iontophoretic pipette in the external plexiform layer was facilitated by monitoring the focal extracellular field potentials with this pipette while it was being advanced through the tissue. The pipettes contained 1 M GABA (pH 4), 1 M sodium glutamate (pH 8) or 1 M sodium aspartate (pH 8).

Stimulation of the olfactory nerves or the mitral cell axons results in a complex of field potentials that can be recorded in the olfactory bulb. The largest wave recorded in the granule cell layer represents the excitation of granule cells¹²⁻¹⁴. Such stimulation results in a prolonged inhibition of the same wave evoked by a second identical stimulus. This paired pulse inhibition is an extracellular measure of the inhibitory postsynaptic potentials (i.p.s.ps) produced in a population of mitral cells that result from the activation of the dendrodendritic reciprocal synaptic pathway^{13,14}. In all five experiments the addition of noradrenaline to the superfusate decreased this paired pulse inhibition (Fig. 1a), indicating that activation of the dendrodendritic inhibitory pathway was, at some point, decreased. This effect was more pronounced at moderate and high stimulus strengths (Fig. 1b). Noradrenaline also depressed the amplitude of the field potentials and therefore, in Fig. 1a, the stimulus strength was increased to evoke a conditioning response identical to that in the control.

It has been demonstrated with intracellular recording from mitral cells that the i.p.s.ps are GABA mediated, chloride dependent¹⁶⁻¹⁸ and can be evoked by a single action potential in a mitral cell^{16,17}. In the present experiments in control conditions, a number of mitral cells displayed successive spontaneous action potentials with intervening i.p.s.ps (Fig. 1c, top trace) which controlled the spontaneous action potential frequency. Thus, if the preparation was not stimulated, spikes occurred at a very regular interval which was determined by the amplitude and rate of decay of the i.p.s.ps. When the superfusate was switched to one containing noradrenaline (5–100 μ M), both the i.p.s.p. following the spontaneous action potentials and the olfactory nerve-evoked i.p.s.ps (marked with arrows) were invariably shortened (Fig. 1c, bottom trace; $n = 11$). In addition, the peak conductance increase associated with the i.p.s.ps was decreased. The i.p.s.ps following spontaneous action potentials were also decreased in amplitude. In those cells which were not spontaneously firing action potentials noradrenaline had no effect on the membrane potential or conductance, although in some cells the threshold for spike initiation increased. This last effect might account in part for the depressant action of noradrenaline on the field potentials (see above) and the inhibitory effects reported *in vivo*²².

The diminution of the i.p.s.p. could be due to the action of noradrenaline on either mitral or granule cells or both. As neither resting membrane potential nor membrane resistance of mitral cells was affected by noradrenaline, its site of action might be on the granule cells, although it could exert its effects

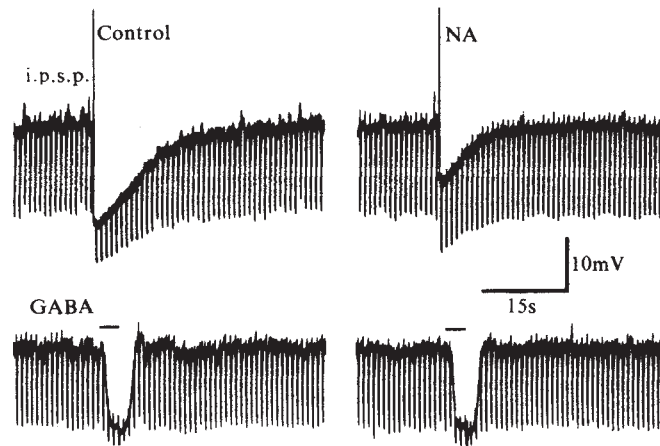


Fig. 3 Noradrenaline is not a GABA antagonist. The records on the left are the control i.p.s.p. (top) and control GABA response (bottom). The records on the right were obtained 24 min after the addition of noradrenaline (20 μ M) to the superfusate. GABA was applied with a current of 80 nA. Resting potential was -54 mV throughout.

by blocking calcium entry (see refs 23, 24) into mitral cells. To examine this question further, the effects of noradrenaline were tested on mitral cell responses in the presence of tetrodotoxin. The calcium spikes that can be elicited in the presence of tetrodotoxin¹⁵⁻¹⁷ were not reduced by concentrations of noradrenaline up to 100 μ M while the i.p.s.p. was reduced substantially (Fig. 2; $n = 3$). Furthermore, this could be observed in cells in which spike threshold was unaltered.

Next, by iontophoresing GABA onto mitral cells we tested the possibility that noradrenaline might attenuate i.p.s.ps by blocking the action of the inhibitory transmitter. However, concentrations of noradrenaline that markedly reduced the size of the i.p.s.p. did not affect the iontophoretic GABA response (Fig. 3; $n = 4$). Thus, we conclude that the primary site of action of noradrenaline is on the granule cells. It has been reported that noradrenaline can enhance the action of GABA in the cerebellum⁵. We did not see such an enhancement, although small increases could have gone undetected. However, as shown above the overall effect of noradrenaline on synaptic inhibition in the olfactory bulb is depression.

In contrast to the actions of noradrenaline, iontophoretic glutamate enhanced the amplitude and duration of i.p.s.ps evoked by spontaneous action potentials (Fig. 4a; $n = 3$). This prolongation decreased the frequency of spikes without a change in spike threshold. When the iontophoretic current was increased, complete cessation of spontaneous action potentials occurred concomitant with a hyperpolarization of the membrane potential and a decrease in membrane resistance (Fig. 4b). Very short but intense ejections of glutamate resulted in fast hyperpolarizations associated with decreases in membrane resistance. Both effects outlasted the iontophoretic pulse and took the same form as i.p.s.ps (Fig. 4c). This hyperpolarizing action of glutamate could be observed in the presence of tetrodotoxin (Fig. 4d). However, after the addition of the GABA antagonist, bicuculline (Fig. 4d), or after blockade of synaptic transmission by cobalt, glutamate evoked a depolarization ($n = 8$). This indicates that the direct action of glutamate on mitral cells is a depolarization and suggests that, like the action of noradrenaline, glutamate acts on mitral cells primarily by modulating the release of GABA from granule cells.

Our observations suggest that noradrenaline and also enkephalin²⁵ decrease the i.p.s.p. recorded in mitral cells by acting on granule cells, causing them to release less GABA in response to the same stimulus. Glutamate and aspartate exert the opposite effect on granule cells. Based on extracellular recording techniques, McLennan²⁶ suggested that glutamate and noradrenaline might be acting in a similar way on granule

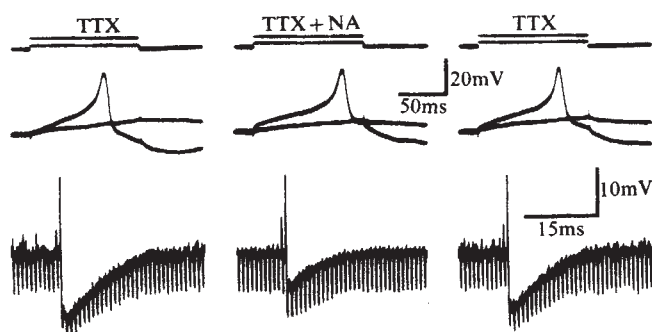


Fig. 2 Diminution of the i.p.s.p. in the presence of tetrodotoxin by noradrenaline. Top row: oscilloscope records of calcium spikes in tetrodotoxin (1 μ M), after 15 min of superfusion with 100 μ M noradrenaline, and after 23 min of washing in tetrodotoxin. Bottom row: simultaneous pen records of i.p.s.ps following calcium spikes. Resting potential was -60 mV throughout. The voltage calibration bar represent 1 nA for the current trace.

cells. The reason for this discrepancy is unclear. Noradrenaline may produce its disinhibitory effect by shunting or hyperpolarizing the granule cell membrane, thus attenuating excitatory input, or by interfering with excitation-release coupling^{23,27}. Although a direct effect on the release process in mitral cells has not been completely ruled out, mitral cell calcium spikes are unaffected by noradrenaline, and thus any interference with release would presumably have to occur after calcium entry.

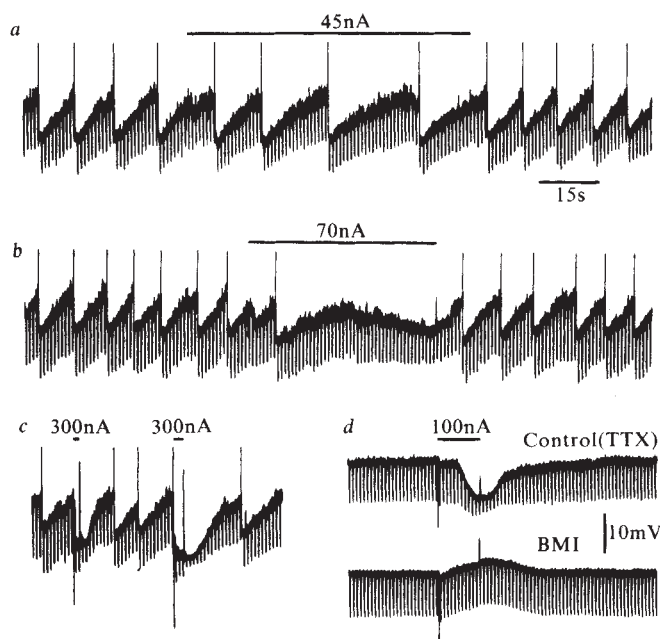


Fig. 4 Effect of glutamate on mitral cell i.p.s.ps and membrane potential. *a-c*, Ionophoretic applications of glutamate onto the same cell. Spike threshold was -57 mV throughout. *d*, Glutamate iontophoresis in a preparation perfused with tetrodotoxin ($1 \mu\text{M}$) before (top trace) and 20 min after the addition of 0.1 mM bicuculline methiodide (BMI) (bottom trace). Resting potential was -55 mV.

The hyperpolarizing action of glutamate and aspartate on mitral cells has been shown to result from the release of GABA from granule cells. This is presumably due to a depolarizing action of the amino acids on granule cells. The fact that doses of excitatory amino acid below that which caused the release of GABA still enhanced the evoked release of GABA suggests that subthreshold depolarization of granule cells facilitates excitation-secretion coupling. This result agrees with recent findings in invertebrate neurones²⁸.

In conclusion, it is proposed that noradrenaline, enkephalin and glutamate alter mitral cell excitability primarily by indirect actions; noradrenaline and enkephalin block GABA release from granule cells, whereas glutamate enhances GABA release. The granule cell, then, is the final common pathway on which the centrifugal fibres and their transmitter candidates exert presynaptic control and thereby modulate the flow of information through the olfactory bulb. Such a disinhibitory role for noradrenaline in a sensory pathway fits well into the proposed role of the locus coeruleus noradrenergic system in arousal¹¹, a behavioural state associated with enhanced neuronal responsiveness²⁹.

This research was supported by NIH grant NS 16485, RCDA NS 00287 to R.A.N., and by the Klingenstein Fund.

Received 1 October 1981; accepted 12 March 1982.

1. Moore, R. Y. & Bloom, F. E. *A. Rev. Neurosci.* **2**, 113-168 (1969).

2. Fallon, J. G. & Moore, R. Y. *J. comp. Neurol.* **180**, 533-544 (1978).

3. Morrison, J. H., Molliver, M. E., Grzanna, R. & Coyle, J. T. *Neuroscience* **6**, 139-158 (1981).

4. Bloom, F. E. in *Psychopharmacology—A 20 Year Progress Report* (eds Lipton, M. E., Killam, K. C. & Di Mascio, A.) 131-142 (Raven, New York, 1978).
5. Woodward, D. J., Moises, H. C., Waterhouse, B. D., Hoffer, B. J. & Freedman, R. *Fedn. Proc.* **38**, 2109-2116 (1979).
6. Segal, M. *Brain Res.* **206**, 107-128 (1981).
7. Langmoen, I. A., Segal, M. & Andersen, P. *Brain Res.* **208**, 349-362 (1981).
8. Nakai, Y. & Takaori, S. *Brain Res.* **71**, 47-60 (1974).
9. Rogawski, M. A. & Aghajanian, G. K. *Nature* **287**, 731-734 (1980).
10. Mueller, A. L., Hoffer, B. J. & Dunwiddie, T. V. *Brain Res.* **214**, 113-126 (1981).
11. Aston-Jones, G. & Bloom, F. E. *J. Neurosci.* **1**, 887-900 (1981).
12. Rall, W. & Shepherd, G. M. J. *Neurophysiol.* **31**, 884-915 (1968).
13. Nicoll, R. A. *Brain Res.* **14**, 157-172 (1969).
14. Mori, K. & Takagi, S. F. *J. Physiol., Lond.* **305**, 171-196 (1980).
15. Mori, K., Nowicky, M. C. & Shepherd, G. M. J. *J. Physiol., Lond.* **314**, 295-309 (1981).
16. Jahr, C. E. & Nicoll, R. A. *Science* **207**, 1473-1475 (1980).
17. Jahr, C. E. & Nicoll, R. A. *J. Physiol., Lond.* (in the press).
18. Nowicky, M. C., Mori, K. & Shepherd, G. M. J. *Neurophysiol.* **46**, 639-648 (1981).
19. Collins, G. S. & Probert, G. A. *Brain Res.* **209**, 231-234 (1981).
20. Price, J. L. & Powell, T. P. S. *J. Cell Sci.* **7**, 157-187 (1970).
21. Halasz, N., Ljungdahl, A. & Hokfelt, T. *Brain Res.* **154**, 253-271 (1978).
22. Bloom, F. E., Costa, E. & Salmoiraghi, G. C. *J. Pharm. exp. Ther.* **146**, 16-23 (1964).
23. Dunlap, K. & Fischbach, G. D. *J. Physiol., Lond.* **317**, 519-535 (1981).
24. Horn, J. P. & MacAfee, D. A. *J. Physiol., Lond.* **301**, 191-204 (1980).
25. Nicoll, R. A., Alger, B. E. & Jahr, C. E. *Nature* **287**, 22-25 (1980).
26. McLennan, H. *Brain Res.* **29**, 177-184 (1971).
27. Shapiro, E., Castellucci, V. F. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1185-1189 (1980).
28. Shapiro, E., Castellucci, V. F. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 629-633 (1980).
29. Livingstone, M. S. & Hubel, D. H. *Nature* **291**, 554-561 (1981).

Differential release of serotonin and histamine from mast cells

Theoharis C. Theoharides, Philip K. Bondy*, Nikolaos D. Tsakalos† & Philip W. Askenase

Department of Internal Medicine, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA and * Veterans Administration Medical Center, West Spring Street, West Haven, Connecticut 06516, USA

Exocytosis dependent on calcium and metabolic energy has been established as the mechanism for the release of membrane-bound secretory products from various exocrine, endocrine and neural cells¹. This has also been shown to be the case in mast cells, which have been used increasingly as a model secretory system². The secretory granules of mast cells contain several mediators³, some of which, such as histamine, are known to participate in many immune reactions and allergic diseases^{4,5}. Because of mast cell involvement in these clinical syndromes, as well as the role of histamine in gastric acid secretion⁶ and possibly in brain pathophysiology⁷, there has been great interest in the pharmacological modulation of histamine release from mast cells⁸. Serotonin is also stored in mast cell granules of several species but much less is known about its secretion. Because histamine and serotonin may have divergent functions in delayed hypersensitivity^{4,9}, we hypothesized that these amines could undergo differential release. We now report that the tricyclic antidepressant drug amitriptyline (Elavil) inhibits histamine release from stimulated mast cells while permitting the release of serotonin. In these conditions, exocytosis of secretory granules is largely prevented, but serotonin is released by an unknown process which still requires calcium and metabolic energy. The ability to secrete differentially expands the physiological potential of the mast cell, and suggests that release of serotonin may not always indicate mast cell secretion via exocytosis of secretory granules.

Rat peritoneal mast cells were collected, purified (90% purity) and incubated as described previously¹⁰. Mast cells that had been preincubated with ³H-serotonin to load them with this releasable amine¹¹ and then incubated with amitriptyline (10^{-7} – 10^{-4} M) for 5 min showed depressed histamine release

† Permanent address: Department of Internal Medicine, Aristotelian University School of Medicine, Thessaloniki, Greece.

Table 1 Amitriptyline permits differential release from mast cells

Amitriptyline (M)	Mast cell release (% total)					
	Histamine		Serotonin		Ruthenium red staining	
	48/80	IgE + Ag	48/80	IgE + Ag	48/80	IgE + Ag
—	71 ± 3	40 ± 9	83 ± 4	46 ± 6	83 ± 6	52 ± 1
10 ⁻⁷	70 ± 4	38 ± 7	83 ± 5	45 ± 8	82 ± 7	50 ± 5
10 ⁻⁶	65 ± 5	30 ± 8	78 ± 6	38 ± 6	79 ± 5	41 ± 6
5 × 10 ⁻⁶	59 ± 6	21 ± 5	75 ± 3	37 ± 7	68 ± 6	35 ± 4
10 ⁻⁵	40 ± 3	17 ± 6	74 ± 4	39 ± 5	57 ± 4	28 ± 3
5 × 10 ⁻⁵	27 ± 3	12 ± 4	72 ± 3	34 ± 3	34 ± 3	13 ± 3
10 ⁻⁴	11 ± 2	7 ± 4	76 ± 5	39 ± 2	19 ± 3	9 ± 1
Inhibition of secretion (% total)						
10 ⁻⁴	85	83	8	15	77	83

Effect of amitriptyline on mast cell release of histamine and serotonin. Mast cells (90% purity, 10⁶ cells ml⁻¹) were first incubated in HEPES-buffered Locke's solution¹⁰ with 5 μ Ci ml⁻¹ ³H-serotonin (29.8 Ci mmol⁻¹; NEN) for 1 h at 37 °C, and were then incubated (10⁵ cells per tube) with 0–10⁻⁴ M amitriptyline for 5 min, followed by an additional 5 min incubation with compound 48/80 (1 μ g ml⁻¹) or 30 min incubation with anti-ovalbumin hybridoma IgE antibody (1:1,000 dilution) and ovalbumin (Ag, 100 μ g ml⁻¹) in the presence of phosphatidylserine (100 μ M). The histamine released was assayed fluorometrically³⁸ while the serotonin released was measured by liquid scintillation counting¹¹, both in duplicate portions. Ruthenium red (0.005%) staining of mast cells was performed as described previously¹⁴. Spontaneous release was 1.8% for serotonin, 3.6% for histamine. Results represent the mean $\pm$ s.d. of five separate determinations.

but nearly unaltered serotonin release when stimulated either nonspecifically with the classic mast cell secretagogue¹² compound 48/80 (1 μ g ml⁻¹) or immunologically by addition of 1:1,000 dilution of mouse ascites fluid from an IgE antibody-secreting hybridoma¹³ and antigen (ovalbumin, 100 μ g ml⁻¹) in the presence of 100 μ M phosphatidylserine (Table 1). When experiments were performed with unpurified peritoneal cells (10% mast cells), higher concentrations of amitriptyline were required to permit differential release. Exocytosis of secretory granules was identified by selective staining with ruthenium red which, by virtue of its relative inability to cross the plasma membrane and its high affinity for proteoglycans of the secretory granule core, binds to and stains only those granules exposed to the extracellular fluid¹⁴. This assay showed that depression of histamine release by amitriptyline was accompanied by a substantial reduction in apparent granule exocytosis (Table 1). To examine the requirements necessary for the release of serotonin but not histamine, mast cells were preincubated with either EDTA (5 × 10⁻³ M for 2 h) or antimycin A (10⁻⁶ M) and 2-deoxyglucose (5 × 10⁻⁴ M) for 30 min followed by amitriptyline and then compound 48/80, or IgE and antigen as before. With calcium depletion or deprivation of metabolic energy, serotonin failed to appear in the supernatant (Table 2). Thus, differential release of serotonin seemed to be calcium and energy dependent.

Although histamine release has been the traditional measure of mast cell secretion, release of radiolabelled serotonin—presumably incorporated into mast cell granules from the extracellular medium—has been used increasingly as an alternative means of quantifying mast cell secretion because of its simplicity and sensitivity, and because it has so far been shown

to parallel the release of histamine^{11,15–17}. However, our results indicate that serotonin can be released in the absence of substantial histamine secretion, and in the absence of demonstrable exocytosis of secretory granules. To exclude the possibility that differential release was solely due to an unknown ability of exogenous radiolabelled serotonin to bind to some location from which it could be displaced without exocytosis of secretory granules, we used a HPLC system to assay endogenous (pre-formed and granule-stored) serotonin both in supernatants and in cell pellets of mast cells stimulated to secrete by compound 48/80 in the presence of amitriptyline (10⁻⁴ M). In conditions where there was only ~11% release of histamine and 19% staining by ruthenium red (Table 2), there was ~67% release of endogenous serotonin. These results show that activation of mast cells in the presence of amitriptyline resulted in similar release of both exogenous and endogenous serotonin.

Data from three previous studies suggest that serotonin may indeed be secreted differentially from mast cells^{18–20}. Other studies have reported that different mast cell secretory products may be packaged either in distinct granules or in other cytoplasmic components from which they may be mobilized separately^{21–22}. However, as serotonin and histamine are both thought to be stored in the secretory granules of mast cells, it is necessary to postulate a mechanism whereby serotonin, and not histamine, could be transported from the granules to the cell exterior. It is possible that in some circumstances serotonin may be preferentially moved from the secretory granules to another compartment, perhaps in small cytoplasmic vesicles from which it may then be secreted by energy- and calcium-requiring exocytosis. There is some evidence that serotonin is present in the cytoplasm of mast cells^{23,24}, and a transport system for serotonin that involves a specific serotonin binding protein released by exocytosis together with serotonin has been demonstrated in the nervous system²⁵. Furthermore, a vesicular transport mechanism has recently been identified in guinea pig basophils²⁶, and we have identified proteins in rat basophil leukaemia cells and mast cells that bind serotonin²⁷. We suggest that specific serotonin-binding proteins, perhaps in cytoplasmic vesicles, may be involved in differential release by mast cells.

The ability of amitriptyline to inhibit histamine release (which has also been shown in human basophils²⁸) and to permit differential release of serotonin, may be partly related to its ability to inhibit phosphatidylinositol turnover²⁹, which has been linked to the cascade of events involved in mast cell secretion³⁰. This differential release contrasts with the explosive release of many mast cell granule constituents that is commonly seen in allergic and immediate hypersensitivity disease mechanisms, and is often used to raise doubts about the physiological usefulness of such a dramatic reaction^{31,32}.

Our present finding, that there may be a way of triggering the selective release of mast cell mediators, is also supported by recent *in vivo* work. In these experiments, immunized mice were challenged with antigen and developed skin swelling responses associated with mast cell degranulation, release of serotonin and formation of gaps between endothelial cells³³. However, mice treated with pargyline—another drug permitting differential secretion from mast cells³⁴—showed no mast cell degranulation but had skin swelling and evidence of

Table 2 Energy and calcium dependence of differential release of serotonin

Stimulus	Mast cell secretion (% total)											
	No inhibitory drug						Amitriptyline (10 ⁻⁴ M)					
	Histamine		Serotonin		Trypan blue staining		Histamine		Serotonin		No	
	—	No Ca ²⁺	No energy	—	No Ca ²⁺	No energy	—	No Ca ²⁺	No energy	—	No Ca ²⁺	No energy
48/80	70 ± 3	7 ± 2	6 ± 2	81 ± 5	5 ± 2	8 ± 3	12 ± 5	4 ± 1	7 ± 2	81 ± 7	9 ± 2	8 ± 3
IgE + Ag	38 ± 9	6 ± 3	5 ± 1	36 ± 8	9 ± 3	7 ± 2	9 ± 6	5 ± 2	4 ± 1	42 ± 6	7 ± 2	5 ± 2
											2 ± 1	3 ± 1
											1 ± 1	2 ± 1
												3 ± 1

Effect of energy or calcium deprivation on differential release of serotonin in the presence of amitriptyline (10⁻⁴M) and in response to either compound 48/80 (1 μ g ml⁻¹) or IgE antibody plus appropriate antigen as described in Table 1 legend. Mast cell viability was tested with the Trypan blue (0.04%) exclusion method and was >94%. Results represent the mean $\pm$ s.d. of five separate determinations.

serotonin-induced gaps between endothelial cells³⁵. These results, with the recently demonstrated ability of mast cells to secrete individual granules in well localized areas of their plasma membrane^{36,37}, indicate that mast cells could participate in

previously unrecognized functions, some of which may be involved in syndromes that are alleviated by antidepressant agents.

This work was supported in part by USPHS and NIH grants AI-12211, AI-11077 and AI-10497 and AI-17555.

Received 7 December 1981; accepted 2 March 1982.

1. Douglas, W. W. *Ciba Fdn Symp.* **54**, 61–90 (1978).
2. Garland, L. G. & Mongar, J. L. *Symp. Soc. exp. Biol.* **30**, 193–218 (1976).
3. Ho, P. C., Lewis, R. A., Austen, K. F. & Orange, R. P. *Comprehensive Immun.* **6**, 179–228 (1979).
4. Askenase, P. W. *Prog. Allergy* **23**, 199–320 (1977).
5. Kazimierzczak, W. & Diamant, B. *Prog. Allergy* **24**, 295–365 (1978).
6. Sell, A. & Walsh, J. H. A. *Rev. Physiol.* **41**, 35–53 (1979).
7. Green, J. P., Maayani, S., Weinstein, H. & Hough, L. B. *Psychopharmac. Bull.* **16**, 36–38 (1980).
8. Foreman, J. C. & Lichtenstein, L. M. A. *Rev. Med.* **31**, 181–190 (1980).
9. Gershon, R. K., Askenase, P. W. & Gershon, M. D. *J. exp. Med.* **142**, 732–747 (1975).
10. Theoharides, T. C. & Douglas, W. W. *Endocrinology* **102**, 1637–1640 (1978).
11. Morrison, P. C., Roser, J. F., Henson, P. M. & Cochrane, C. G. *J. Immun.* **112**, 573–582 (1975).
12. Röhlich, R., Anderson, P. & Uvnäs, B. *J. Cell Biol.* **51**, 465–483 (1971).
13. Böttcher, J., Hammering, G. & Kapp, J. F. *Nature* **275**, 761–762 (1978).
14. Nemeth, E. F. & Douglas, W. W. *Naunyn-Schmiedeberg's Archs Pharmac.* **302**, 153–163 (1978).
15. Otsuki, J. A., Grassick, R., Seymour, D. & Kind, L. S. *Immun. Commun.* **5**, 27–39 (1976).
16. Mazine, C., Dessaint, J.-P. & Capron, A. *J. immun. Meth.* **21**, 65–77 (1978).
17. Chasin, M., Scott, C., Shaw, C. & Persico, F. *Int. Archs Allergy appl. Immun.* **58**, 1–10 (1979).

18. Goldstein, D. J., Finkelman, S. & Nahmond, V. E. *Medicine* **34**, 584–585 (1974).
19. Miller, P. & Church, M. K. *Int. Archs Allergy appl. Immun.* **52**, 53–58 (1976).
20. Ichikawa, A., Kaneko, H., Mori, Y. & Tomita, K. *Biochem. Pharmac.* **26**, 197–202 (1977).
21. Lynch, S. M., Austen, K. F. & Wasserman, S. F. *J. Immun.* **121**, 1394–1399 (1978).
22. Sannes, P. S. & Spicer, S. S. *Am. J. Path.* **94**, 447–456 (1979).
23. Calsson, S.-A. & Ritzen, M. *Acta physiol. scand.* **77**, 449–464 (1969).
24. Gustafsson, B. *Int. Archs Allergy appl. Immun.* **63**, 121–128 (1980).
25. Tamir, H. & Gershon, M. D. *J. Neurochem.* **33**, 35–44 (1979).
26. Dvorak, A. M. *et al. Lab. Invest.* **42**, 263–276 (1980).
27. Tamir, H., Theoharides, T. C., Gershon, M. D. & Askenase, P. W. *J. Cell Biol.* (in the press).
28. Lichtenstein, L. M. & Gillepsie, E. J. *Pharmac. exp. Ther.* **192**, 441–450 (1975).
29. Smith, T. L. & Hauser, G. *Biochem. Pharmac.* **28**, 1759–1763 (1979).
30. Cockcroft, S. & Gomperts, B. D. *Biochem. J.* **178**, 681–687 (1979).
31. Padawer, J. *Am. J. Anat.* **141**, 299–302 (1974).
32. Kaliner, M. A. *New Engl. J. Med.* **501**, 498–500 (1979).
33. Askenase, P. W., Bursztajn, S., Gershon, M. D. & Gershon, R. K. *J. exp. Med.* **152**, 1358–1374 (1980).
34. Askenase, P. W. & Theoharides, T. C. *Fedn Proc.* **39**, 905 (1980).
35. Askenase, P. W., Schwartz, A., Siegel, J. N. & Gershon, R. K. *Int. Archs Allergy appl. Immun.* **66** (Suppl. 1) 225–233 (1981).
36. Lawson, D., Fewtrell, C. & Raff, M. C. *J. Cell Biol.* **79**, 394–400 (1978).
37. Theoharides, T. C. & Douglas, W. W. *Science* **201**, 1143–1145 (1978).
38. Kremzner, L. T. & Wilson, I. B. *Biochim. biophys. Acta* **50**, 364–367 (1961).

H-2K-, H-2I- and H-2D-restricted hybridoma contact sensitivity effector cells

M. Minami, K. Okuda, M. E. Sunday & M. E. Dorf

Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA

Mice primed with 4-hydroxy-3-nitrophenylacetyl-*O*-succinimide (NP-*O*-Su) generate hapten-specific T cell-mediated contact or cutaneous hypersensitivity (CS) responses¹. Analysis of the effector T cells mediating these *in vivo* responses suggested that multiple T-cell subpopulations were involved². It has not been possible previously to separate these cell populations physically by conventional means. Therefore, to separate these cells and to provide a uniform source of CS effector cells for future analyses, we hybridized the CS effector cell population with AKR-derived BW5147 thymoma cells. We now describe five H-2-restricted hybridoma T-cell clones isolated from this fusion. Each hybridoma possesses hapten-specific CS activity and four also cross-react with 4-hydroxy-3-nitrophenylacetyl (NP) derivatives. Two of the hybridoma clones are specific for NP associated with H-2I gene products; two others for NP associated with H-2D products; and one for NP associated with H-2K gene products.

C57BL/6 mice were immunized with 2 mg NP-*O*-Su; 6 days later T cells were prepared from the regional lymph nodes as

previously described for the enrichment of Ts₃ suppressor cells³. The hybridization procedure has been reported elsewhere³, in fact the same series of hybridoma cells were used to screen for CS and suppressor cell activity. The hybridoma colonies were screened for CS reactivity by injecting 10⁵ irradiated cells (2,000 R) together with either 40 µg of NP-bovine serum albumin (BSA) or 5 × 10⁵ NP-coupled C57BL/6 spleen cells into the left footpad. Maximal antigen-specific swelling responses occurred 24–48 h after injection.

The specificity of five CS reactive hybridoma clones is demonstrated in Table 1. Two types of NP-reactive hybridoma cell lines were distinguished. The B6-CSI-15 and B6-CSI-78 lines demonstrated significant footpad swelling responses when given together with NP-coupled proteins (for example, NP-bovine γ-globulin, NP-BGG), but failed to respond when transferred with carrier protein (BGG) alone. Similar NP-specific swelling responses were noted when the hybridoma cells were injected with other NP-coupled proteins, including NP-BSA (Table 3) and NP-keyhole limpet haemocyanin (KLH) (data not shown). However, these hybridoma lines failed to give NP-specific responses when administered with NP-coupled syngeneic C57BL/6 spleen cells. In contrast, the B6-CSK-21, B6-CSD-35 and B6-CSD-52 hybrids demonstrated NP-specific swelling responses when transferred with NP-coupled C57BL/6 spleen cells, but failed to respond to trinitrophenyl (TNP)-coupled C57BL/6 spleen cells or NP-coupled proteins. Significant levels of footpad swelling were elicited after local injection of as few as 10³ hybridoma cells plus antigen. Injections of 10⁴ or 10⁵ hybridoma cells resulted in larger swelling responses, while injection of 10⁶ cells gave swelling responses only slightly larger than those induced by 10⁵ cells. In contrast, when 10³–10⁶

Table 1 Specificity of hybridoma CS effector cells

Hybridoma cell line	Challenge antigen					
	NP-BGG	NIP-BGG	BGG	NP-cells	NIP-cells	TNP-cells
BW5147	0.3 ± 0.9	1.0 ± 0.6	0.3 ± 0.3	1.5 ± 0.6	NT	NT
B6-CSI-15	8.3 ± 1.7*	6.0 ± 0.6*	0.6 ± 0.3	0.8 ± 0.6	NT	NT
B6-CSI-78	8.3 ± 1.5*	7.3 ± 1.5*	0.3 ± 0.3	1.5 ± 0.9	NT	NT
BW5147	0.9 ± 1.3	NT	NT	0.3 ± 0.7	0.3 ± 0.3	1.3 ± 0.7
B6-CSK-21	0.9 ± 0.6	NT	NT	8.0 ± 0.6*	10.0 ± 0.6*	0.3 ± 0.3
B6-CSD-35	0.9 ± 0.9	NT	NT	8.3 ± 0.3*	0.6 ± 0.3	0.0 ± 0.6
B6-CSD-52	1.3 ± 0.6	NT	NT	8.3 ± 0.3*	8.7 ± 0.9*	1.3 ± 0.9

1 × 10⁵ irradiated BW5147 cells or hybridoma cells were mixed with 40 µg NP-BGG, NIP-BGG, BGG, or 5 × 10⁵ NP-, NIP-, or TNP-conjugated C57BL/6 spleen cells, and these were then injected into the left footpad of C57BL/6 recipients in 25 µl Hanks' balanced salt solution. Footpad swelling, measured as the difference in footpad thickness between the left and right footpads 24 h after antigenic challenge, is recorded in units of 10⁻³ cm. The groups each contain 4–5 mice. NT, not tested.

* Significant swelling response ($P < 0.01$).

Table 2 Phenotypes of hybridoma CS effector cells

Hybridoma cell line	Antibody specificity							Mouse immunoglobulin
	Thy-1.1	Thy-1.2	H-2D ^b	I-A ^b	I-J ^b	NP ^b	CGAT	
BW5147	98	7	3	5	0	2	0	1
B6-CSI-15	98	98	84	0	3	24	2	0
B6-CSI-78	98	98	74	3	2	4	3	0
B6-CSK-21	98	98	74	7	2	3	2	2
B6-CSD-35	100	98	73	6	2	2	1	0
B6-CSD-52	98	7	78	5	2	2	0	0
B6-Ts ₁ -3	98	5	84	0	53	31	0	0

Monoclonal hybridoma anti-Thy-1.1, -Thy-1.2 (NEN) and -I-A^b antibodies (Dr T. Springer, Sidney Farber Cancer Institute, Massachusetts), conventional anti-H-2D^b ((A × 18R)F₁ anti-B10) and anti-I-J^b (5R anti-3R) alloantisera, guinea pig anti-NP^b and anti-CGAT idiotypes, and polyvalent guinea pig anti-mouse immunoglobulin antisera (prepared as detailed elsewhere⁹) were used, together with predetermined optimal concentrations of rabbit complement, to lyse the hybridoma cells. The results are expressed as per cent specific lysis and represent pooled results from 2–6 experiments. Less than 15% lysis is considered insignificant. The B6-Ts₁-3 T-cell hybridoma line was used as a control. Characterization of monoclonal Ts₁ suppressor cells is detailed elsewhere⁶. CGAT is the common idotype on anti-GAT antibodies¹⁰.

hybridoma cells were given intravenously (i.v.) and the mice were challenged with antigen in the footpads, no swelling response was observed (data not shown). This may be due to the failure of the hybridoma cells to properly migrate to the site of antigenic challenge. However, note that the use of a local transfer system has limitations: as helper⁴ and killer^{4,5} T cells can also induce delayed inflammatory responses, it is difficult to determine whether these hybridoma cells are CS effector cells or represent hybridomas derived from distinct T-cell subsets. This may be a semantic problem².

Previous reports^{1,2} have indicated that NP-specific CS effector cells derived from Igh^b-bearing C57BL/6 mice cross-reacted with NP derivatives such as 4-hydroxy-5-iodo-3-nitrophenylacetyl (NIP). To evaluate the fine specificity of the CS reactive hybridoma clones isolated, they were injected with either NIP-derivatized proteins or cells. As shown in Table 1, four of the five hybridomas demonstrated significant levels of NIP cross-reactivity. Previous analyses of the C57BL/6-derived NP-O-Su-induced CS effector cell population did not reveal a non-NIP-cross-reactive population^{1,2} but these previous experiments involved adoptive transfers of a heterogeneous population of H-2-restricted cells. Thus the presence of some NIP-cross-reactive clones in the heterogeneous population probably prevented detection of other non-cross-reactive clones.

To further distinguish between these hybridoma clones, the cells were phenotyped by microcytotoxicity testing⁶ using antisera reactive with Thy-1, H-2- and Igh-V-encoded determinants derived from the C57BL/6 parental strain. The AKR (H-2^k, Igh^d)-derived BW5147 thymoma cells and the NP-specific suppressor T-cell hybridoma line B6-Ts₁-3 of C57BL/6 (H-2^b, Igh^b) origin, were used as controls. All the lines carried the BW5147-derived Thy 1.1 marker (Table 2), as shown by the high percentage of lysis observed after treatment with monoclonal anti-Thy-1.1 antibody and complement. Most of the hybridoma clones also carried the Thy-1.2 marker derived from the spleen cell donors. The B6-CSD-52 and B6-Ts₁-3 lines were exceptions—this may be due to the loss of the chromosome

carrying the Thy-1.2 gene^{7,8}. All the C57BL/6-derived hybridomas specifically reacted with anti-H-2D^b antisera. However, none of the hybridomas demonstrated significant reactivity with monoclonal anti-I-A^b or conventional anti-I-J^b antibodies⁹. The cytolytic activity and specificity of the anti-I-J^b antisera described previously⁶ were confirmed by testing the I-J^b-bearing B6-Ts₁-3 hybridoma suppressor line.

To evaluate the nature of the antigen receptor on these hybridomas, the cells were tested for reactivity with guinea pig anti-NP^b idiotype antisera⁶. Control guinea pig anti-idiotype antisera directed against a different idiotype determinant associated with anti-GAT (poly(Glu⁶⁰Ala³⁰Tyr¹⁰)) antibodies¹⁰ and polyvalent guinea pig anti-immunoglobulin antisera were used as controls. Only the B6-CSI-15 and the control B6-Ts₁-3 hybridoma lines demonstrated specific lysis with anti-NP^b antisera. All cloned sublines of B6-CSI-15 also demonstrated <50% lysis with anti-NP^b antisera, suggesting that the idiotype receptor is poorly expressed. Similar findings were reported for the idiotype receptors on suppressor cell hybridomas¹¹. In addition, note that only one of the four hybridoma lines which cross-reacted with NIP hapten (Table 1) carried detectable levels of NP^b-related idiotype determinants, indicating that at the T-cell level, NIP cross-reactivity does not correlate with the presence of NP^b idiotype determinants. These results are also consistent with data which indicate that B6-derived CS effector cells are generally not susceptible to complement-mediated lysis with anti-NP^b antisera (unpublished data).

The hybridoma CS effector lines were further distinguished by evaluating the genetic restrictions of activation. The B6-CSI-15 and B6-CSI-78 hybridoma lines only demonstrated NP-specific reactivity when transferred into I-A^b-bearing recipients (Table 3). Thus, transfer of B6-CSI-15 and B6-CSI-78 cells together with NP-BSA yielded significant footpad swelling responses in I-A^b-bearing C57BL/6 and B10.A(5R) recipients, but not in strains lacking an I-A subregion derived from the H-2^b haplotype. In contrast, when B6-CSK-21 cells were transferred together with NP-coupled syngeneic spleen cells, footpad

Table 3 Genetic restriction of hybridoma CS effector cells

Hybridoma cell line	Challenge antigen	Recipient strains					
		B10.A(5R) bbbbbddd	B10.MBR bkkkkkkq	B10.A(4R) kkbbbbb	B10.A(2R) ddddd	B10.A kkkkkkd	C57BL/6 bbbbbbb
BW5147	NP-BSA	1.0 ± 0.4	1.0 ± 0.4	1.0 ± 0.7	NT	1.0 ± 0.4	1.3 ± 0.6
B6-CSI-15	NP-BSA	7.3 ± 1.4*	1.0 ± 0.9	0.3 ± 1.3	NT	1.0 ± 0.7	15.0 ± 1.6*
B6-CSI-78	NP-BSA	9.8 ± 1.4*	1.3 ± 0.6	1.8 ± 0.9	NT	1.5 ± 0.6	10.5 ± 1.6*
BW5147	NP-cells	0.8 ± 0.5	0.8 ± 0.3	NT	0.8 ± 0.5	1.0 ± 0.4	0.3 ± 0.3
B6-CSK-21	NP-cells	10.8 ± 2.3*	9.8 ± 1.4*	NT	0.3 ± 0.5	1.3 ± 0.9	9.8 ± 0.8*
B6-CSD-35	NP-cells	0.5 ± 0.6	0.5 ± 0.6	NT	10.3 ± 0.9*	1.0 ± 0.4	17.0 ± 1.8*
B6-CSD-52	NP-cells	2.5 ± 1.0	0.5 ± 1.0	NT	8.0 ± 1.1*	0.8 ± 0.6	9.0 ± 1.0*

See Table 1 legend for protocol. Recipients received 5×10^5 syngeneic hapten-modified spleen cells.

swelling responses were noted only in C57BL/6, B10.A(5R) and B10.MBR recipients. Thus, the activity of the B6-CSK-21 hybridoma cell line is restricted by genes in the *H-2K* region. Finally, the C57BL/6-derived B6-CSD-35 and B6-CSD-52 hybrids demonstrated NP-specific responses only in *H-2D*- (and *H-2I*-) compatible C57BL/6 and B10.A(2R) recipients.

The present series of *H-2K*-, *H-2I*- and *H-2D*-restricted T-cell hybrids are specific for the same haptenic determinants. Although all the hybrids are NP-specific, the B6-CSI lines do not react with NP-coupled cells and the B6-CSK and B6-CSD lines fail to respond to NP-coupled proteins. This may reflect the interaction requirements for activation of *H-2I*-restricted hybridomas by cells which process and present antigen, while the *H-2D*- or *H-2K*-restricted cells may be activated by a different kind of antigen association^{2,12,13}. These data support the hypothesis that activation of the *H-2K*- and *H-2D*-

restricted effector CS hybridoma cells is macrophage-independent, as are the interactions of cloned or hybridoma cytolytic T effector cells¹⁴⁻¹⁶.

The series of NP-specific hybridomas isolated and described here provides a very useful tool for analysing antigen receptors on T cells. Further analysis of these hybridoma lines may help to elucidate T-cell function, including analysis of the signals required for cellular activation; comparisons of the products released after antigen activation and determining whether these cells mediate multiple T-cell functions (for example, whether *H-2I*-restricted hybrids may provide helper activity or whether the *H-2K/D*-restricted cells function as killer T cells). Experiments are being done to address these issues.

We thank the Cancer Research Institute and the NIH (AI-16677 and CA-14732) for support, and T. Dinse and N. Axelrod for secretarial assistance.

Received 26 October 1981; accepted 25 March 1982.

1. Sunday, M. E., Weinberger, J. Z., Benacerraf, B. & Dorf, M. E. *J. Immun.* **125**, 1601-1605 (1980).
2. Sunday, M. E. & Dorf, M. E. *J. Immun.* **127**, 766-768 (1981).
3. Okuda, K., Minami, M., Furusawa, S. & Dorf, M. E. *J. exp. Med.* **154**, 1838-1851 (1981).
4. Dennert, G., Weiss, S. & Warner, J. F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4540-4543 (1981).
5. Lin, Y.-L. & Askonas, B. A. *J. exp. Med.* **154**, 225-234 (1981).
6. Okuda, K., Minami, M., Ju, S.-T. & Dorf, M. E. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4557-4561 (1981).
7. Marcucci, F., Waller, M., Kirchner, H. & Krammer, P. *Nature* **291**, 79-81 (1981).

8. Minami, M., Okuda, K., Furusawa, S., Benacerraf, B. & Dorf, M. E. *J. exp. Med.* **154**, 1390-1402 (1981).
9. Bhattacharya, A., Dorf, M. E. & Springer, T. A. *J. Immun.* **127**, 2488-2495 (1981).
10. Ju, S.-T., Benacerraf, B. & Dorf, M. E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6192-6196 (1978).
11. Kanno, M., Takei, I., Suzuki, N., Tomioka, H. & Taniguchi, M. *Proc. jap. Soc. Immun.* **10**, 41-42 (1980).
12. Leung, K. N. & Ada, G. L. *Scand. J. Immun.* **12**, 481-487 (1980).
13. Ertl, H. C. J. *Cell. Immun.* **62**, 38-49 (1981).
14. Gillis, S. & Smith, K. A. *Nature* **268**, 154-156 (1977).
15. Reiss, C. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 5432-5436 (1980).
16. Kaufmann, Y., Berke, G. & Eshhar, Z. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2502-2506 (1981).

Failure of killed *Listeria monocytogenes* vaccine to produce protective immunity

Carl Heinz Wirsing von Koenig & Horst Finger

Institute of Hygiene and Laboratory Medicine, Municipal Hospital, D-4150 Krefeld, FRG

Herbert Hof

Institute of Hygiene and Microbiology, University of Wuerzburg, D-8700 Wuerzburg, FRG

We report here experiments which, in contrast to recently published data^{1,2}, suggest that protection against lethal doses of viable *Listeria monocytogenes* cannot be achieved by previous injection with killed *Listeria* vaccine neither by adding macrophage-blocking agents such as dextran sulphate nor by using C3H/HeJ mice having inborn macrophage defects. This discrepancy can be explained by the ineffective killing of bacteria at 56 °C, resulting in injection of small numbers of viable bacteria; such doses are normally subinfective and do not provide immunity. Blocking of the macrophage system by dextran sulphate³, however, or use of C3H/HeJ mice with inborn macrophage defects⁴ renders these doses fully infective, hence providing subsequent protective immunity⁵.

L. monocytogenes, serotype 4b, was cultured in tryptose broth at 37 °C for 16 h. The virulence of the strain was maintained by continuous passage in mice of the same strain as used for the experiments. Viable counts of bacteria in broth, dilutions and spleen homogenates were determined using 10-fold dilutions on to agar. Dilutions, where necessary, were made in sterile 0.9% phosphate-buffered saline (PBS) pH 7.2.

Using the protocol of van der Meer *et al.*², the bacterial suspensions were incubated at 56 °C for 60 min to kill the bacteria. Then the suspensions were diluted to a final concentration of 5.0×10^8 bacteria ml⁻¹. In the first experiment, 0.2 ml of this dilution (1.0×10^8 bacteria) was injected intraperitoneally (i.p.) into female NMRI (Han) mice (Central Institute for Laboratory Animals, Hannover), aged 8-12 weeks. The animals were divided into two groups; one group received an additional i.p. injection of 50 mg dextran sulphate 500

(DS 500) per kg body weight 1 day before infection (Pharmacia). Viable counts of the bacterial suspensions revealed in five independent experiments, between 7.5×10^1 and 8.0×10^2 *L. monocytogenes* per ml, resulting in infective doses between 1.5×10^1 and 1.6×10^2 bacteria per mouse. The time dependence of bacterial killing at 56 °C is shown in Table 1.

The time course of infection of mice was followed by counting bacteria in spleen homogenates at different times after infection. As can be seen from Fig. 1, 1.6×10^2 bacteria per mouse did not result in detectable viable bacteria in the spleens, without further treatment. Challenge of these animals 1 week later with an i.p. injection of $20 \times \text{LD}_{50}$ (lethal dose for 50% of the population, 4.5×10^5) *L. monocytogenes* led to the death of all animals during the subsequent 5-day observation period. Figure 1 shows that the second group of mice, which were treated with the macrophage blocking agent DS 500, became infected after i.p. injection of only 1.6×10^2 bacteria. The course of infection resembled the normal course of experimental listeriosis⁶. Subsequent challenge of these animals with $20 \times \text{LD}_{50}$ of viable *L. monocytogenes* revealed protective immunity as only one of eight animals died within the period of observation.

The second experiment used C3H/HeJ and C3HeB/FeJ, male mice (24-28 g; Jackson Laboratories), which have an 'intrinsic adjuvant' (ref. 1). Earlier studies have shown that C3H mice belong to a group of mouse strains that are susceptible to infection with *L. monocytogenes*, a phenomenon thought

Table 1 Killing of *Listeria monocytogenes* at 56 °C

Incubation time (min)	Viable bacteria per ml
0	3.5×10^9
30	2.8×10^6
40	2.8×10^2
50	8.0×10^1
60	1.2×10^2
70	2.0×10^1
80	0
90	4.0×10^1
100	0

L. monocytogenes serotype 4b were grown at 37 °C for 16 h in tryptose broth. The broth was incubated in a 56 °C waterbath then chilled to 4 °C before the viable count was performed.

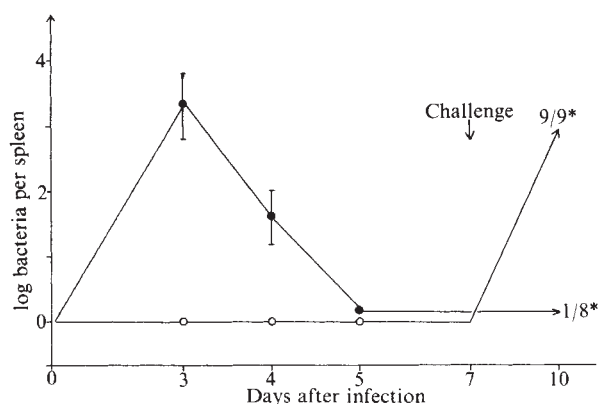


Fig. 1 Course of infection with *Listeria monocytogenes* in DS 500-treated mice. Female NMRI mice were injected i.p. with 1.6×10^2 viable *L. monocytogenes*. ○, Animals given no additional treatment; ●, animals injected i.p. with 50 mg DS 500 per kg body weight on day 1 before infection. Seven days later, all mice were challenged with $20 \times \text{LD}_{50}$ (4.5×10^5) viable *L. monocytogenes* i.p. Each point represents the mean $\pm$ s.e.m. of five spleens. * Proportion of animals dead at end of observation period.

to be mediated by a single autosomal gene, *Lr*⁷. The HeJ substrain of C3H mice has a macrophage defect⁴ which renders it highly susceptible to facultative intracellular parasites such as *Salmonella typhimurium*⁸. Consequently, the i.p. LD_{50} was 4.0×10^3 *L. monocytogenes* compared with 2.3×10^4 for C3HeB/FeJ mice. Both mouse strains were infected with a suspension of bacteria produced as described above. The doses of viable bacteria were fully infective for C3H/HeJ but not for C3HeB/FeJ mice, resulting in induction of protective immunity in the former in the same manner as that shown for DS-500-treated NMRI mice.

We then repeated the experiments using injections of *L. monocytogenes* killed by incubation at 70 °C for 60 min. The killed bacteria were unable to produce immunity or to increase the nonspecific resistance to infection in both DS-500-treated NMRI and in C3H/HeJ mice. We therefore conclude that blocking or impairment of the mononuclear phagocyte system by either DS 500 or *Bordetella pertussis* organisms^{9,10}, or by inborn defects as in the C3H/HeJ mouse substrain, renders normally subinfective doses of *L. monocytogenes* fully infective, which can be clearly demonstrated by a drastically reduced mean lethal dose. Animals surviving this infection, however, possess a strong protective immunity based on the interaction of T-lymphocytes and macrophages.

These experiments stress the importance of a functionally active macrophage system for resistance against facultative intracellular parasites. Contrary to the findings of van Dijk¹ and van der Meer², our experiments suggest that injection with a killed *Listeria* vaccine is unable either to provide protective immunity or to increase nonspecific resistance to infection.

This work was supported by Bundesministerium des Innern.

Directed effector cells selectively lyse human tumour cells

Charles B. Simone

Clinical Pharmacology Branch, Division of Cancer Treatment, National Cancer Institute, Bethesda, Maryland 20205, USA

Classical antibody-dependent cellular cytotoxicity (ADCC) involves effector cells that mediate damage of antibody-coated targets. Classical ADCC is probably responsible, at least in part, for *in vivo* allograft rejections, resistance to viral infections, parasite destruction and the rejection of tumours^{1,2}. Like complement^{3,4}, human effector cells mediate damage by creating a pore or channel in the target membrane, but the channel created by effector cells is 2.5 times larger (functional diameter $\sim 165 \text{ \AA}$) than the channel formed by complement^{3,5}. Classical ADCC is inhibited by low concentrations of immune complexes which compete with target-bound antibody for effector cell Fc receptors⁶. Because many clinically important diseases, such as certain cancers and systemic infections, have been shown to result in circulating antigen-antibody complexes, it seems likely that these immune complexes might inhibit the ADCC component of host defense mechanism in these diseases. A major advantage to the host might result from the attachment of antibody to effector cells first; either via the Fc receptor⁷, which is a weak interaction; by antibody 'associating' with effector cells⁸, which is very transient; or by binding antibody to the cellular membrane of effector cells⁹. Here I report the specificity of such antibody-directed ADCC effectors for antigen-bearing targets *in vitro*.

Antibody can be attached to effector cells by incubating them together in the presence of polyethylene glycol (PEG) and spinning them through a mixture of phthalate oils (see Table 1 legend). Using radiolabelled antibody, it can be shown that 150,000 antibody molecules can be attached to each effector cell (neutrophil) by this technique (data not shown). Also, no radiolabelled antibody is detected for up to 24 h in supernatants of directed effectors incubated at 37 °C; or for up to 6 h in the plasma of mice (five animals) which were injected intravenously with directed lymphocytes (data not shown). The effects of

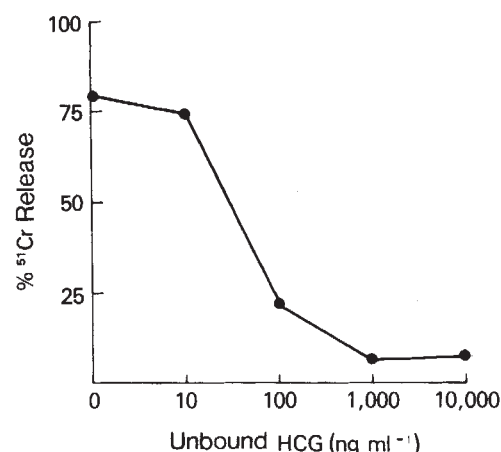


Fig. 1 Effect of unbound HCG on directed effector cell action on ⁵¹Cr-labelled JEG-3 cells. 7.5×10^5 ⁵¹Cr-labelled -3 cells in RPMI 1640 medium were suspended with 5% fetal calf serum in flat-bottom microtitre wells with indicated amounts of unbound HCG. Anti-HCG-directed effector cells were introduced into the well to a total volume of 200 μ l and the plate was gently rotated for 10 min, then incubated for 5 hr at 37 °C. ⁵¹Cr release was assayed as before. Controls were effector cells treated with PEG and phthalate oils without antibody, which yielded 7–10% ⁵¹Cr release at all points of unbound HCG.

Received 3 December 1981; accepted 26 March 1982.

- van Dijk, H., Hofhuis, F. M. A., Berns, E. M. J. J., van der Meer, C. & Willers, J. M. N. *Nature* **286**, 713–714 (1980).
- van der Meer, C., Hofhuis, F. M. A. & Willers, J. M. N. *Nature* **269**, 594–595 (1977).
- Hahn, H. & Bierther, M. *Infect. Immunity* **10**, 110–119 (1974).
- Rosenstreich, D. L. & Vogel, S. N. in *Microbiology* 1980, 11–15 (ASM, Washington, D.C., 1980).
- Hasenclever, H. F. & Karakawa, W. W. *J. Bact.* **74**, 584–586 (1957).
- Mackness, G. B. *J. exp. Med.* **116**, 381–406 (1962).
- Skamene, E., Konghavi, P. A. L. & Sachs, D. H. *J. infect. Dis.* **139**, 228–231 (1979).
- O'Brien, A. D. *et al. J. Immun.* **124**, 20–24 (1980).
- Hof, H., Emmerling, P., Finger, H. & Wirsing, C. H. *Zentbl. Bakt. Hyg. I. Abt.* **A240**, 208–214 (1978).
- Finger, H., Heymer, B., Wirsing, C. H., Emmerling, P. & Hof, H. *Infect. Immunity* **19**, 950–960 (1978).

Table 1 Action of anti-TNP-directed effector cells on TNP-modified targets

Effector	Anti-TNP (1 µg ml ⁻¹)	% Marker release		
		TNP-ghosts (CF)	TNP- ⁵¹ Cr- RBC	TNP- ⁵¹ Cr- MCF-7
2 × 10 ⁶ PBLs	0	11	8	10
1 × 10 ⁶ Neutrophils	0	13	7	7
2 × 10 ⁶ PBLs*	0	13	9	10
1 × 10 ⁶ Neutrophils*	0	12	11	11
Directed ADCC				
2 × 10 ⁶ PBLs	+	94	70	67
1 × 10 ⁶ Neutrophils	+	101	77	76
Classical ADCC				
2 × 10 ⁶ PBLs	+	96	68	68
1 × 10 ⁶ Neutrophils	+	93	72	76

Human resealed erythrocyte ghosts containing CF were made as described earlier³. Intact RBCs and MCF-7 cells were TNP-modified and ⁵¹Cr-labelled. Samples were counted in a γ-counter. Monolayers of each target were made on the bottom surface of flat-bottom microtitre wells as previously described³. Freshly separated human granulocytes and lymphocytes were obtained as described previously³ and used for classical and directed ADCC. They were washed three times and resuspended in balanced salt solution (BSS)-HEPES to a final concentration of 5 × 10⁷ cells per ml. 'Directed' ADCC effector cells were prepared as follows. Affinity-purified rabbit anti-TNP IgG was mixed with effector cells. 12% of 20,000-molecular weight PEG (Fisher Scientific) in BSS-HEPES was immediately added, bringing the total mixture to 1 mg ml⁻¹ immunoglobulin in 6% PEG. The mixture was incubated for 90 min at 0°C and then 250 µl aliquots were layered over 800 µl of a phthalate oil mixture (dibutyl phthalate:bis-(2-ethylhexyl)phthalate, both from Eastman Chemicals, Rochester; 1.5:1 v/v) in 1.5 ml Beckman microcentrifuge tubes and sedimented in a Beckman Microfuge B at 9,360 g for 4 min. Cells can be separated from the incubation medium without an aqueous wash because the oil mixture is less dense than the cells and more dense than the medium. The supernatant and entire oil mixture were recovered by aspirating them from the tubes. The cells were washed five times using BSS-HEPES and resuspended to the original concentration of 5 × 10⁷ cells per ml. Control effector cells* were treated with PEG and the phthalate oils as above but without antibody. The microtitre wells were all refilled with BSS-HEPES and the desired reaction components (effector cell, antibody) were added to a final volume of 200 µl. The plates were centrifuged at 250 g for 10 min and incubated at 37°C for 3 h; then 100 µl aliquots were withdrawn from each well and fluorescence or ⁵¹Cr release determined. The total releasable marker was determined by the marker released by 1% Triton X-100. Samples were run in duplicate and agreement between the duplicates was within 15%.

classical ADCC were compared with those of 'directed' ADCC (Table 1) using trinitrophenyl (TNP)-modified targets. Peripheral blood lymphocytes (PBLs) or neutrophils were added to anti-TNP-coated targets (classical ADCC), or effector cells with attached anti-TNP antibody ('directed' ADCC) were added to TNP labelled targets. The three targets used were TNP-resealed human erythrocyte ghosts containing carboxyfluorescein (CF), TNP-⁵¹Cr-labelled intact human erythrocytes and TNP-⁵¹Cr-labelled MCF-7 cells, a human breast cancer cell line¹¹; each was put in separate microtitre wells. Effector cells alone or effector cells treated with PEG and the phthalate oils but without antibody produced minimal background release from each target, showing that the PEG and phthalate oil treatment *per se* does not cause target damage. Directed effector cells can mediate marker release from a membrane model (erythrocyte ghosts), intact erythrocytes and even from human breast cancer cell MCF-7. These results are comparable with those seen when classical ADCC methods are used. Hence the attachment of antibody to effector cells using this technique does not inhibit cell-mediated killing.

The specificity of the directed cytotoxic effector cells was assayed by placing two antigenically distinct targets in the same microtitre well: TNP-modified ghosts containing CF and human chorioncarcinoma cell line, JEG-3 (gift of Dr Saul Rosen), labelled with ⁵¹Cr. The JEG-3 cell line secretes human chorionic gonadotropin (HCG) at a rate of 7,290 IU per 10⁹ cells per day; hormone secretion can, however be blocked by antibody against HCG¹². If anti-TNP-directed-effector cells are used, very high CF release from the TNP-modified ghost population is observed and there is minimal background ⁵¹Cr release from the ⁵¹Cr-labelled JEG-3 cells (see Table 2). However, if anti-

HCG-directed effector cells (anti-HCG antibody given by Dr Hans Hager, Hoffmann-LaRoche) are used, then ⁵¹Cr release is high and CF release is minimal, indicating that the JEG-3 cells are killed and the TNP-ghosts are not damaged. The specificity of directed ADCC is comparable with that of classical ADCC. When soluble immune complexes (HCG-anti-HCG) were added to microtitre wells containing 10⁶ JEG-3 cells before 10⁶ anti-HCG-directed neutrophils, there was 36% ⁵¹Cr net release after 4 h of incubation compared with 40% net release using the same system but without immune complexes. This was in contrast to only 7% ⁵¹Cr net release from 10⁶ JEG-3 cells which were treated with anti-HCG first, then with immune complexes, then 10⁶ effector neutrophils. Hence immune complexes do not significantly inhibit directed ADCC but do inhibit classical ADCC.

JEG-3 cancer cells were chosen as targets because the anti-HCG-directed effector cell-JEG 3 system may be a model clinically relevant to human testicular cancer. It is of interest, therefore, to determine whether unbound HCG (analogous to circulating HCG in a patient with testicular cancer) interferes with the lysis of JEG-3 cancer cells mediated by anti-HCG-directed effector cells. Figure 1 shows that ≥1,000 ng ml⁻¹ unbound HCG almost totally prevents anti-HCG-directed effectors from lysing ⁵¹Cr-labelled JEG-3 cancer cells, probably by binding to the anti-HCG antibody attached to the effector cells. Whereas 100 ng ml⁻¹ of unbound HCG allows partial ⁵¹Cr release, 0 or 10 ng ml⁻¹ of unbound HCG does not interfere with killing.

Recently, Miller *et al*¹³ treated a T-cell leukaemia patient with systemic anti-Leu 1 antibody and a clinical response, albeit transient was demonstrated. Several difficulties encountered in that patient, including clearance of antibody and dosage, might be avoided in future human *in vivo* therapy if the technique described here was used. Hence the anti-human Leu 1 (Becton Dickinson) reagent was used to see whether it would effect lysis of targets *in vitro* after it had been attached to effector cells in the conditions already described (Table 3). Freshly separated PBLs were ⁵¹Cr-labelled and divided into two groups, one of which was TNP-modified. When anti-TNP-directed effectors were put into either group, there was significant ⁵¹Cr release from the TNP-modified PBLs and only background ⁵¹Cr from the unmodified PBL targets. When anti-human Leu 1-directed

Table 2 Directed compared with classical ADCC using anti-TNP or anti-HCG and their action on targets

Effectors	Anti-TNP		Anti-HCG	
	TNP-ghosts % CF release	JEG 3 cells % ⁵¹ Cr release	TNP-Ghosts % CF release	JEG 3 cells % ⁵¹ Cr release
Directed ADCC				
2 × 10 ⁶ PBLs				
+ antibody	83	8	6	47
1 × 10 ⁶ Neutrophils				
+ antibody	92	7	7	53
Classical ADCC				
2 × 10 ⁶ PBLs				
+ antibody	92	7	8	49
1 × 10 ⁶ Neutrophils				
+ antibody	90	8	8	54
No antibody				
	TNP-Ghosts	JEG cells		
2 × 10 ⁶ PBLs	11	7		
1 × 10 ⁶ Neutrophils	11	7		
2 × 10 ⁶ PBLs*	13	8		
1 × 10 ⁶ Neutrophils*	14	8		

4 × 10⁵ TNP-modified erythrocyte ghosts containing CF and 4 × 10⁵ ⁵¹Cr-labelled JEG-3 cells were suspended together in flat-bottom microtitre wells in RPMI 1640 medium. Desired components were then added and marker release assayed after 5 h incubation at 37°C. Directed effector cells were made as before and control effector cells* were treated with PEG and phthalate oils without antibody. Anti-HCG is affinity-purified rabbit IgG.

Table 3 ADCC effector cells directed with anti-TNP or anti-Leu 1 antibodies and their effects on TNP-⁵¹Cr-labelled PBL or ⁵¹Cr-labelled PBL targets

Effector cell Treatment	% ⁵¹ Cr release	
	TNP- ⁵¹ Cr-PBLs	⁵¹ Cr-PBLs
Directed ADCC		
1 × 10 ⁶ Neutrophils*	15	11
1 × 10 ⁶ Neutrophils + anti-TNP	74	12
1 × 10 ⁶ Neutrophils + anti-Leu 1	61	56
Classical ADCC		
1 × 10 ⁶ Neutrophils*	10	13
1 × 10 ⁶ Neutrophils + anti-TNP	71	13
1 × 10 ⁶ Neutrophils + anti-Leu 1	59	62

Freshly separated PBLs were ⁵¹Cr-labelled and divided into two groups, one of which was TNP-modified. 7.5 × 10⁵ PBLs of each group were put in separate microtitre wells and suspended in 0.1 M BSS-HEPES and desired components were added to a total volume of 200 µl. Directed effectors were made as before and other control effectors* were treated with PEG and phthalate oils but without antibody. After incubation for 5 h at 37 °C, 100 µl supernatant aliquots were removed and ⁵¹Cr content assayed.

effector cells were used, ⁵¹Cr was released from each group, but slightly less than the amount seen for the anti-TNP-directed effector cell TNP-PBL reaction. Presumably this is because only the T-cell subpopulation is attacked when anti-human Leu 1 reagent is used, whereas T cells, B cells and null cells are attacked in the TNP-PBL group when anti-TNP is used. Classical ADCC results are comparable.

The technique of using PEG and phthalate oils to attach antibody to freshly separated effector cells results in the specific lysis of cells bearing the target antigen. The lysis is specifically blocked by soluble antigen. Soluble immune complexes do not inhibit directed ADCC. The degree of specific target lysis of human cancer cells and model targets using these directed effector cells is comparable with the lysis produced with classical ADCC techniques. The phthalate oils are completely washed off directed effector cells and these cells cause no adverse reactions *in vivo* (mice and one human). With the advent of monoclonal antibodies, this technique makes it theoretically possible to cause the destruction of targets *in vivo* (tumours, infectious agents) provided the antibody-antigen system is well defined. The absolute amount of non-human antibody given to a patient using this technique (and hence the risk of serum sickness and anaphylactic reactions) should be much less than that amount of antibody passively administered to patients with certain tumours and Gram-negative sepsis¹⁴. Trials using passive administration of non-human antibody in an attempt to produce human tumour regression have been limited^{13,15-17} because recent evidence shows that cell-mediated rather than humoral immunity is important in tumour resistance^{18,19} and that passive antibody may enhance tumour growth²⁰. I have shown here that human effector cells can be selectively directed to lyse human cancer cells and not other adjacent cells.

I thank Drs A. S. Fauci, D. Longo, C. Myers and F. Abramson for helpful comments.

Received 4 November 1981; accepted 26 March 1982.

- Cerottini, J. C. & Brunner, K. T. *Adv. Immun.* **18**, 67 (1974).
- Waksman, B. H. in *Mechanisms of Lymphocyte Activation*, 249 (Wiley, New York, 1974).
- Simone, C. B. & Henkart, P. A. *J. Immun.* **124**, 954 (1980).
- Giavedoni, E. B., Chow, Y. M. & Dalmasso, A. P. *J. Immun.* **122**, 240 (1979).
- Dourmashkin, R. R., Deteix, P., Simone, C. B. & Henkart, P. A. *Clin. exp. Immun.* **42**, 554 (1980).
- MacLennan, I. C. M. *Clin. exp. Immun.* **10**, 275 (1972).
- Segal, D. M. & Hurwitz, E. *J. Immun.* **4**, 1338 (1977).
- Imir, T., Sakela, E. & Makela, O. *J. Immun.* **117**, 1938 (1976).
- Jones, J. F. & Segal, D. F. *J. Immun.* **125**, 926 (1980).
- Greenburg, A. H. & Shen, L. *Nature* **245**, 282 (1973).
- Soule, H. D., Vasquez, E. R., Long, A., Alberts, S. & Brennan, M. J. *J. natn. Cancer Inst.* **51**, 1409 (1973).
- Kohler, P. O. *et al. Acta endocr. Suppl.* **150-154**, 137 (1971).
- Miller, R. A., Maloney, D. G., McKillop, J. & Levy, R. *Blood* **58**, 78 (1981).
- Ziegler, E. J., McCutchan, J. A. & Braude, A. I. *Trans. Ass. Am. Physns* **91**, 253 (1978).
- Currie, G. A. *Int. J. Cancer* **26**, 141 (1972).
- Wright, P. W., Hellström, K. E., Hellström, I. & Bernstein, I. D. *Med. Clin. N. Am.* **60**, 607 (1976).
- Nadler, L. M. *et al. Cancer Res.* **40**, 3147 (1980).
- Brunner, K. T., Mavel, J., Cerottini, J. C. & Chapuis, B. *Immunology* **14**, 181 (1968).
- Hellström, I. & Hellström, K. E. *Cancer* **34**, 1461 (1974).
- Moller, G. *Nature* **204**, 846 (1964).

Kinetics of induction and molecular size of mRNAs encoding human interleukin-2 and γ -interferon

Shimon Efrat, Shlomo Pilo & Raymond Kaempfer

Department of Molecular Virology, The Hebrew University – Hadassah Medical School, 91 010 Jerusalem, Israel

T-cell, growth factor (TCGF), or interleukin-2 (IL-2), and immune or γ -interferon (IFN- γ) are lymphokines possessing powerful immunoregulatory properties. IL-2 is essential for the *in vitro* proliferation and maturation of certain classes of T lymphocyte^{1,2}, particularly cytotoxic^{3,4} and helper⁵ T cells, and also activates natural killer cells⁶. IFN- γ is considerably more active than IFN- α and - β as activator of natural killer cells and as antiproliferative agent, relative to its antiviral activity⁷⁻¹¹. As a step towards understanding the regulation of expression of these proteins, we report here the induction of mRNA species encoding IL-2 and IFN- γ in mitogen-stimulated normal human lymphocytes and their expression in *Xenopus laevis* oocytes. On microinjection, distinct species of mRNA sedimenting at 10–10.5S and 13–13.5S direct the synthesis and secretion of active IL-2 and IFN- γ respectively. A second species of IL-2 mRNA is consistently revealed, sedimenting at 13–13.5S and comprising about 20% of the total IL-2 mRNA activity. During induction, an initial concomitant rise in IL-2 and IFN- γ mRNA activities is followed promptly by a concomitant decline in the rate of their accumulation.

We used human tonsils as a convenient source of large amounts of lymphocytes. Cultivation of these cells in the presence of phytohaemagglutinin (PHA) leads to the induction of IL-2 (Fig. 1a). IL-2 activity is detectable in the medium by 8 h and accumulates at an increasing rate until about 24 h, when a distinct decline in the rate of accumulation is observed. The kinetics of IFN- γ induction by PHA in the same culture (Fig. 1b) are remarkably similar to those of IL-2. Inclusion of the tumour promoter 12-O-tetradecanoyl phorbol 13-acetate (TPA) together with PHA enhances both IL-2 and IFN- γ levels in the medium throughout the induction period, but even with TPA present, the rate of accumulation of IL-2 and IFN- γ begins to decline after 20–24 h.

We isolated mRNA from cells that had been cultured for 12 h with PHA and microinjected it into *X. laevis* oocytes. As detailed in Fig. 2 legend, this mRNA directed extensive synthesis of IL-2. About 80% of the expressed IL-2 activity was recovered from the surrounding medium, showing that the mRNA directed both synthesis of active IL-2 and its secretion from oocytes. The mRNA preparation also directed the synthesis of active IFN- γ , the bulk of which was secreted (see Fig. 2 legend).

Table 1 pH sensitivity of IFN- γ and IL-2 activities in oocyte medium

	Culture medium	10S mRNA	13S mRNA
IFN- γ (U ml ⁻¹)			
pH 7.2	2,560	<40	220
pH 3	<40	<40	<40
IL-2 (c.p.m.)			
pH 7.2	35,500	42,500	5,800
pH 3	29,200	38,400	7,800

Half the incubation medium of groups of 10 oocytes, injected with 10S or 13S mRNA as described for Fig. 3, was dialysed for 16 h against phosphate-buffered saline at pH 3 and then for 2 h against buffer at pH 7.2. The other half of the incubation medium was dialysed for 18 h against buffer at pH 7.2. Aliquots of 24 h-LCM were treated similarly. Assays were performed as described in Fig. 2 legend.

The kinetics of appearance of mRNA for IL-2 and for IFN- γ in lymphocytes stimulated by PHA, as followed by microinjection into oocytes, are illustrated in Fig. 2. Accumulation of IL-2 mRNA activity begins after 4 h and follows a sigmoid

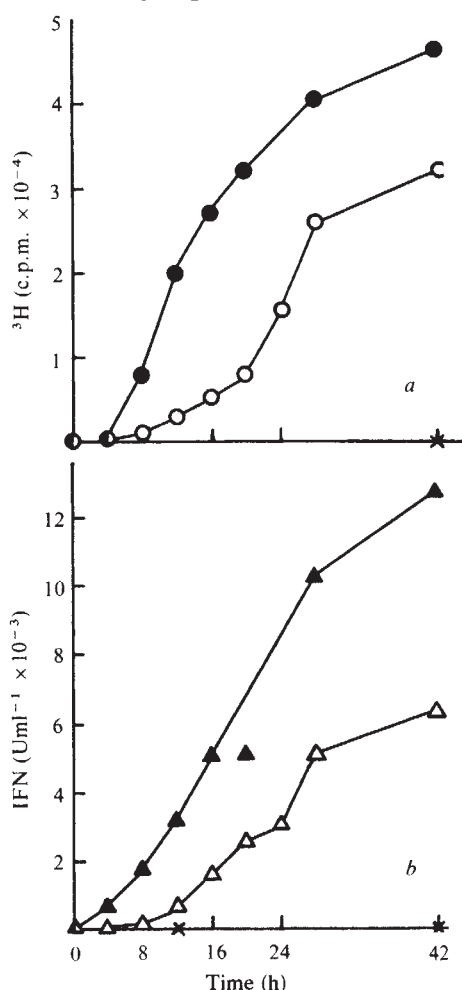


Fig. 1 Kinetics of induction of IL-2 (a) and IFN- γ (b). Tonsils from otherwise healthy donors were disrupted mechanically and cultured at 4×10^6 cells per ml in RPMI 1640 medium supplemented with 2 mM glutamine, 10 mM minimum essential medium (MEM) non-essential amino acids, 100 mM MEM Na-pyruvate, 10 mM HEPES buffer pH 7.2, 5×10^{-5} M 2-mercaptoethanol, 2% fetal calf serum (FCS), antibiotics and 0.4% (v/v) PHA-P (Difco) (○, △). TPA (Sigma) was added to 10 ng ml^{-1} (●, ▲). ×, Mitogen-free controls. At the indicated times, aliquots of the cultures were centrifuged and supernatants were passed through $0.22 \mu\text{m}$ filters and frozen. a, IL-2 activity was assayed by incubating 50 μl of serial dilutions made in culture medium containing 10% FCS with 50 μl of medium containing 2×10^4 assay cells in round-bottom microtitre wells for 48 h at 37°C . Assay cells were human peripheral blood lymphocytes that had been cultured for at least 1 week in IL-2-containing medium and then incubated without IL-2 for 24 h. Each well received $1 \mu\text{Ci}$ of ^3H -thymidine (83 Ci mmol^{-1}) in 10 μl of culture medium at 42 h. The cells were collected on filters and radioactivity in DNA was monitored. Values plotted are for 1:32 dilutions. Values for zero-time cultures, which represent background assay activity of PHA or PHA and TPA, were $<500 \text{ c.p.m.}$ b, IFN- γ was assayed by inhibition of the cytopathic effect of mengovirus in human amnion cells. Confluent cells in microtitre wells were incubated with 50 μl of serial dilutions of the assay fluids in RPMI 1640 medium for 8 h and then infected with 3–5 plaque-forming units (PFU) per cell of mengovirus in 50 μl of medium. After 1 h, the cells were washed and incubated with medium containing 2% FCS for about 24 h before they were stained with Gentian violet. An internal laboratory standard of human IFN- γ and an IFN- α reference standard were included; 1 U of IFN- γ corresponds to about 2.5 IU of IFN- α . The interferon activity was abolished by dialysis against phosphate-buffered saline, pH 2.

course, with a sharp decline in the rate of accumulation at $\sim 16 \text{ h}$. IFN- γ mRNA activity follows a remarkably similar time course (Fig. 2). Comparison with Fig. 1 shows that the levels of mRNA encoding both IL-2 and IFN- γ reach near maximal values well before these activities appear in the medium in significant amounts.

To estimate the molecular size of IL-2 and IFN- γ mRNAs, mRNA extracted from lymphocytes stimulated for 16 h with PHA was centrifuged through an isokinetic sucrose gradient. In Fig. 3a, arrows denote the positions of *Escherichia coli* rRNA and rabbit globin 9S mRNA markers. Microinjection of RNA from individual gradient fractions revealed two peaks of IL-2 mRNA activity, a major one (IL-2A) sedimenting at 10–10.5S and comprising $\sim 80\%$ of the activity, and a minor one sedimenting more rapidly, at 13–13.5S (IL-2B).

Figure 3b depicts a sedimentation velocity analysis of mRNA purified from cells stimulated for 16 h with both PHA and TPA and illustrates the synthesis of both IL-2 and IFN- γ directed by individual gradient fractions. Again, IL-2 mRNA activity sedimented in a major peak at 10S and a minor one at 13S, while IFN- γ mRNA activity sediments in a single peak at 13S that directs high levels of IFN- γ synthesis on microinjection into oocytes. The expression of IFN- γ mRNA in these experiments is at least an order of magnitude higher than that described in previous studies^{12,13} estimating its size at 15S and 18S respectively. IFN- γ encoded by co-sedimenting mRNA might depress DNA synthesis¹¹ in the IL-2 assay cells, in which case the amount of the 13S IL-2 mRNA species would be an underestimate. Alternatively, the overlap of the IL-2B and IFN- γ mRNA peaks at 13S suggests that IFN- γ may induce IL-2 in the cells used for the IL-2 assay. To examine these alternatives, we dialysed the medium of oocytes injected with 10S and 13S mRNA at pH 3 before testing for IL-2 activity. Such treatment effectively eliminates IFN- γ activity but does not affect the IL-2 activity of the two mRNA peaks (Table 1). We conclude, therefore, that the IL-2B mRNA sedimenting at 13S is expressed independently of IFN- γ .

Inclusion of TPA in the culture medium does not influence the relative abundance of the two IL-2 mRNA peaks, but increases by severalfold the specific activities of the IFN- γ mRNA (see Fig. 2) and IL-2 mRNA species (Fig. 3). The higher levels of induction observed when TPA is present (Fig. 1) are thus reflected by increased mRNA activity.

The mRNA species described here were derived from normal human lymphocytes, avoiding abnormal phenomena that may be associated with transformed cell lines¹⁴. The ability of IL-2 to stimulate the proliferation of T cells was assayed in cultured human peripheral blood lymphocytes comprising $>90\%$ T cells, as judged by rosetting with sheep erythrocytes. These cells are IL-2 dependent and die within 48 h when cultured without this factor. At the time of assay, such cells responded only very weakly to PHA or TPA. Although crude IL-2 preparations may contain B-cell growth factor¹⁵ activity in addition to TCGF, the activity observed in our experiments stimulated extensive DNA synthesis in cells that were essentially pure T lymphocytes, thus defining the activity as IL-2.

The expression of human IL-2 mRNA in *X. laevis* oocytes and the secretion of active IL-2 from these cells made it possible to analyse the levels of IL-2 mRNA activity during the course of induction. Mitogenic stimulation of lymphocytes leads to an initial increase in IL-2 mRNA activity that quickly levels off (Fig. 2). The decline in rate of accumulation of IL-2 mRNA activity agrees well with the decline in rate of IL-2 accumulation observed in the culture medium (Fig. 1). It is thus likely that the capacity of an IL-2 producer cell to form active IL-2 mRNA is limited and that a refractory period may follow, similar to that observed for IFN- β induction¹⁶. The kinetics of appearance of both IFN- γ (Fig. 1) and IFN- γ mRNA (Fig. 2) are remarkably similar to those for IL-2. The levelling off of mRNA activities observed in Fig. 2 could be caused by the cessation of mRNA synthesis, by the accumulation of mRNA encoding post-transcriptional repressors, or by both. Gullberg *et al.*¹⁷

have shown that the IL-2-forming capacity of mouse splenocytes declines ~20 h after exposure to mitogen. The present findings support the concept that the induction of IL-2 and IFN- γ in lymphocytes is regulated at the level of transcription.

Finally, our results reveal that IL-2 mRNA is composed of two distinctly sedimenting molecular species (10S and 13S). The mRNA had been treated with dimethyl sulphoxide (DMSO) and heated before centrifugal analysis. This and the observation of a single mRNA peak for IFN- γ , as opposed to two for IL-2, do not support the interpretation that the 13S IL-2 mRNA peak is due to aggregation. This finding probably reflects the existence of more than one gene (or class of genes) encoding IL-2 activity, as has been observed for IFN- β ^{18,19}, although an alternative mRNA processing pathway could also explain the results. The sedimentation coefficient of 10–10.5S for the major IL-2A mRNA species corresponds²⁰ to a nucleotide length of 610–680 bases. Assuming a minimum of 200

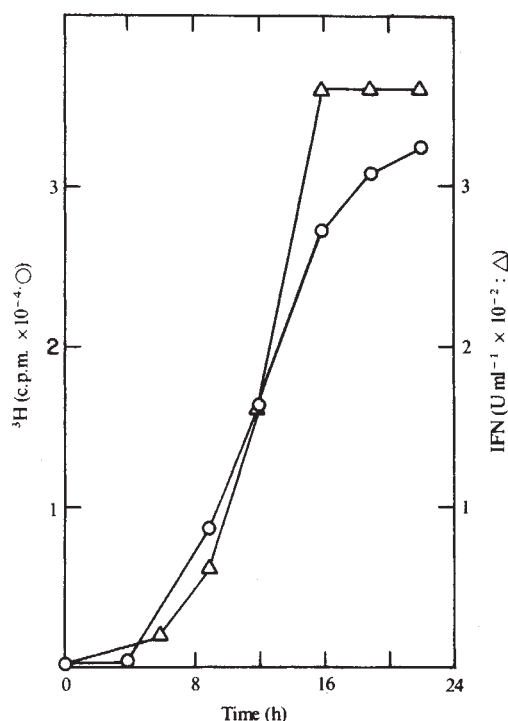


Fig. 2 Kinetics of appearance of IL-2 (○) and IFN- γ (△) mRNA activities in stimulated lymphocytes. Tonsillar lymphocytes were cultured with PHA as described for Fig. 1. At the indicated times, total RNA was extracted by the guanidinium thiocyanate-CsCl method²⁴. The RNA was passed twice over oligo(dT)-cellulose, the second time after incubation for 5 min at 55 °C in 90% DMSO²⁵. Poly(A)-containing RNA was dissolved in water at a concentration of 2 mg ml⁻¹. Oocytes were each injected with 50 nl of mRNA and incubated for 40 h at 21 °C in groups of 10 with 0.1 ml Barth's medium containing 10⁻⁴ M phenylmethylsulphonylfluoride²⁶ (buffer A). Aliquots of incubation medium from each group of oocytes were assayed for IL-2 and IFN- γ activities as described for Fig. 1. Values for IL-2 activity are for 1:4 dilutions. At this dilution, a standard lymphocyte-conditioned medium prepared after 24 h of culturing with PHA (24 h-LCM), gave a count of 30,000 min⁻¹; incubation medium from uninjected oocytes or from oocytes injected with Barth's medium gave <300 c.p.m. background activity. Medium and homogenate (prepared in 0.1 ml of fresh buffer A and cleared by centrifugation) from oocytes each injected with 100 ng of mRNA prepared at 12 h were assayed for IL-2 and yielded 16,400 c.p.m. and 4,400 c.p.m. respectively. When assayed for IFN- γ , these yielded 160 and 25 U ml⁻¹ respectively; uninjected oocytes or oocytes injected with Barth's medium gave no detectable background activity. Interferon activity was abolished by dialysis against pH 2 buffer.

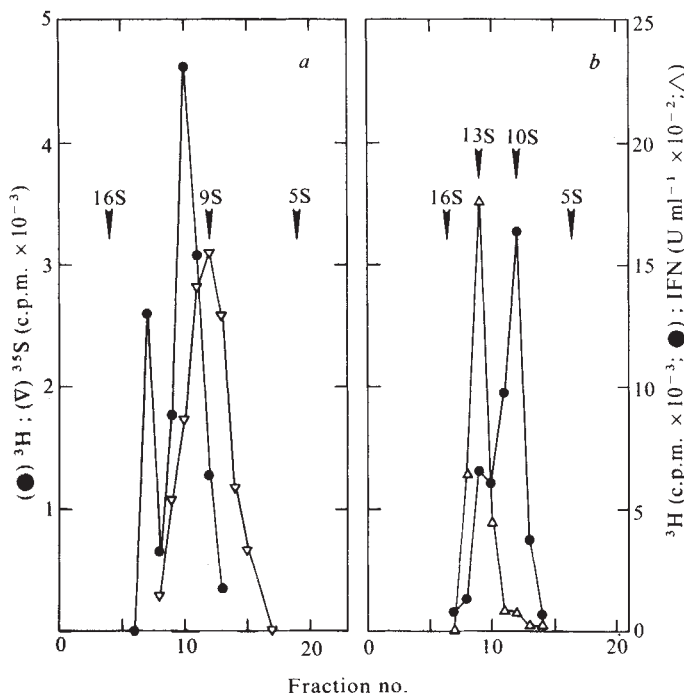


Fig. 3 Sedimentation velocity analysis of IL-2 and IFN- γ mRNAs. Total poly(A)-containing RNA was prepared from lymphocytes cultured for 16 h with PHA (a) or PHA and TPA (b) as described for Figs 1 and 2. The mRNA (160 μ g in a and 80 μ g in b) was heated in water for 90 s at 95 °C, rapidly cooled and centrifuged through 5–17.7% isokinetic sucrose gradients (5 ml) in 10 mM Tris-HCl, pH 7.4, 1 mM EDTA and 0.15 M NaCl. RNA from individual fractions was ethanol-precipitated, dissolved in 10 (a) or 5 (b) μ l of water, and microinjected into 10 oocytes as described for Fig. 2. After incubation, the oocyte medium was assayed for IL-2 (●) and IFN- γ (△), as described for Fig. 1. Values for IL-2 are at a 1:4 dilution; at this dilution, standard 24 h-LCM gave 19,000 c.p.m. Assay backgrounds were as for Fig. 2. The concentration of microinjected RNA in the 10S peak fraction of IL-2 was 0.25 (a) and 0.21 (b) mg ml⁻¹. Parallel gradients contained *Escherichia coli* rRNA A₂₆₀ markers (arrows) or rabbit globin mRNA, monitored for translation activity²⁷ by incorporation of ³⁵S-methionine (V).

nucleotides for the 5'-leader and 3'-untranslated region, including poly(A), the IL-2A mRNA species can thus encode 135–160 amino acids, corresponding to proteins with molecular weights (MWs) of 14,800–17,600 and more if some side chains are modified. This estimate agrees well with the MW of 15,000 reported for IL-2 (refs 21, 22). The minor IL-2B mRNA species and IFN- γ mRNA correspond to a length of 1,060–1,150 nucleotides and accordingly can encode 31,000–35,000 of unmodified protein. These values are in accord with recent estimates of 20–25,000-MW for IFN- γ ²³.

We thank Dr E. Kedar for helpful discussion and assay cells, Dr H. Soreq for advice on microinjection, and S. Marsh, A. Yeheskel and E. Chriqui for assistance. This work was supported by a grant from the Robert Szold Foundation.

Received 31 December 1981; accepted 24 March 1982.

1. Morgan, D. A., Ruscetti, F. W. & Gallo, R. C. *Science* **193**, 1007–1008 (1976).
2. Gillis, S. & Smith, K. A. *Nature* **268**, 154–155 (1977).
3. Gillis, S., Baker, P. B., Ruscetti, F. W. & Smith, K. A. *J. exp. Med.* **148**, 1093–1098 (1978).
4. Zarling, J. M. & Bach, F. H. *Nature* **280**, 685–687 (1979).
5. Tees, R. & Schreier, M. H. *Nature* **283**, 780–782 (1980).
6. Henney, C. S., Kuribayashi, K., Kern, D. E. & Gillis, S. *Nature* **291**, 335–338 (1981).
7. Claeys, H., van Damme, J., de Ley, M., Vermeylen, C. & Billiau, A. *Br. J. Haemat.* **50**, 85–94 (1982).
8. Crane, J. L. Jr, Glasgow, L. A., Kern, E. R. & Younger, J. S. *J. natn. Cancer Inst.* **61**, 871–874 (1978).
9. Sonnenfeld, G., Mandel, A. D. & Merigan, T. C. *Cell. Immun.* **40**, 285–293 (1978).
10. Blalock, J. E., Georgiades, J. A., Langford, M. P. & Johnson, H. M. *Cell. Immun.* **49**, 390–394 (1980).

11. Rubin, B. Y. & Gupta, S. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5928–5932 (1980).
12. Taniguchi, T., Pang, R. H. L., Yip, Y. K., Henriksen, D. & Vilcek, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3469–3472 (1981).
13. Wallace, D. M., Hitchcock, M. J. M., Reber, S. B. & Berger, S. L. *Biochem. biophys. Res. Commun.* **100**, 865–871 (1981).
14. Gillis, S. & Watson, J. J. *exp. Med.* **152**, 1709–1719 (1980).
15. Ford, R. J. *et al. Nature* **294**, 261–263 (1981).
16. Paucker, K. & Boxaca, M. *Bact. Rev.* **31**, 145–184 (1967).
17. Gullberg, M., Ivars, F., Coutinho, A. & Larsson, E. *J. Immun.* **127**, 407–411 (1981).
18. Weissenbach, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7152–7156 (1980).
19. Sehgal, P. B. & Sagar, A. D. *Nature* **288**, 95–97 (1980).
20. Spirin, A. S. *Biokhimiya* **26**, 454–463 (1961).
21. Mier, J. W. & Gallo, R. C. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6134–6138 (1980).
22. Robb, R. J. & Smith, K. A. *Fedn Proc.* **40**, 1163 (1981).
23. Yip, Y. K. *et al. Science* **215**, 411–413 (1982).
24. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294–5299 (1979).
25. Bantle, J. A., Maxwell, I. H. & Hahn, W. E. *Analyt. Biochem.* **72**, 413–427 (1976).
26. Soreq, H. & Miskin, R. *FEBS Lett.* **128**, 305–310 (1981).
27. Di Segni, G., Rosen, H. & Kaempfer, R. *Biochemistry* **18**, 2847–2854 (1979).

The *LSP-2* gene and a 5' flanking sequence are independently expressed in *Drosophila melanogaster*

Barry Yedvobnick* & Michael Levine†

*Department of Biology and

†Department of Molecular Biophysics and Biochemistry,
Yale University, New Haven, Connecticut 06511, USA

The positions of nucleotide sequences involved in the stage- and/or tissue-specific expression of structural genes in higher organisms are largely unknown. One might predict, however, that closely linked genetic loci which are not coordinately controlled will contain regulatory sequences in close proximity to the respective coding regions. During the third instar of larval development in *Drosophila melanogaster*, the hormone ecdysterone augments the expression of several genes in fat body tissue, one of which is the structural gene coding for larval serum protein-2, *LSP-2*^{1–5}. Although *LSP-2* transcripts are detected only in third-instar larval fat bodies, a genomic clone containing the *LSP-2* gene (clone 104) hybridizes to poly(A)⁺ RNA prepared from other tissues. We have now analysed clone 104 for other coding regions. Three transcripts, designated *T1*, *T2* and *T3*, hybridize within the 12.5 kilobases (kb) of DNA flanking the 5' end of the *LSP-2* coding region. The *T1* coding region is separated from the 5' end of the *LSP-2* coding region by no more than 2.5–3.5 kb. In contrast to the dramatic increase in the level of *LSP-2* transcript induced by ecdysterone in third-instar larval fat bodies, the *T1* transcript remains below the detectable level, indicating that expression of the two coding regions is not coordinately regulated.

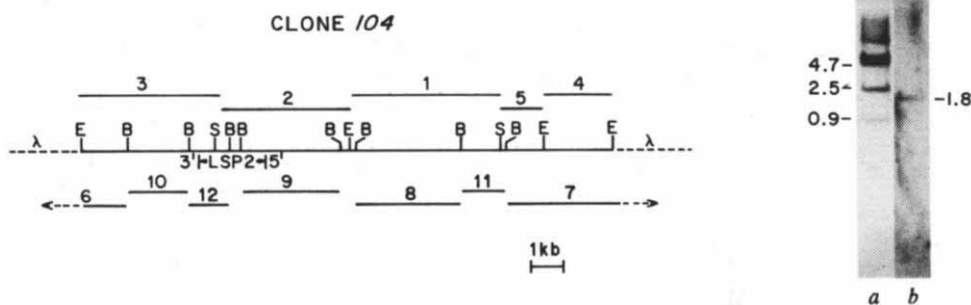
A restriction map for clone 104 (Fig. 1 and ref. 3) shows the cleavage sites of *EcoRI*, *SalI* and *BamHI* endonucleases and the relevant fragments generated by digesting 104 DNA with

either a combination of *EcoRI* and *SalI* (fragments 1–5) or *BamHI* (fragments 6–12). The map location of the 5' terminus of *LSP-2* mRNA was determined by *S*₁ nuclease digestion⁶. The results of this analysis indicate that the 5' end of the *LSP-2* coding region is 1.8 kb from the *SalI* cleavage site in the coding region (Fig. 1). However, we cannot rule out the possibility that the *LSP-2* gene has an intervening sequence which separates a small segment of mRNA coding information from the bulk of the gene. The 1.8 kb value is therefore a minimum estimate.

The initial tests for identifying transcripts that hybridize to 104 DNA involved digestion of the DNA with either a combination of *EcoRI* and *SalI* or *BamHI*, separation of the DNA fragments by electrophoresis in agarose gel, transfer of the fragments from the gel to a sheet of nitrocellulose and hybridization with ³²P-labelled RNA probes (Fig. 2). Intense hybridization to the fragments containing the *LSP-2* coding region (fragments 2, 3, 9, 12) was detected with the fat body RNA probe but not with the oocyte, gastrula or late embryonic RNAs, in agreement with earlier results indicating that the *LSP-2* gene is expressed specifically in third-instar larval fat bodies². Hybridization was also detected with all four RNA probes to fragments in the region flanking the 5' end of the *LSP-2* coding region (fragments 1, 4, 7, 8, 11); fragments 7 and 2 hybridized very weakly with fat body and gastrula RNAs respectively; however, the bands are not visible with the autoradiographic exposures shown in Fig. 2. Hybridization to fragment 11 was visible with the fat body RNA probe with shorter exposures. No hybridization was detected with any of the RNA probes to DNA fragments containing sequences flanking the 3' end of the *LSP-2* coding region.

The transcripts hybridizing with 104 DNA were further characterized by electrophoresis of poly(A)⁺ RNA from late embryos and total RNA from fat bodies in agarose gel, transfer of the fractionated RNA to nitrocellulose paper and hybridization with ³²P-labelled 104 DNA probes (Fig. 3). When intact 104 DNA was used as the probe, a single transcript correspond-

Fig. 1 Restriction map of clone 104. The clone was isolated from a library of Charon-4 phages containing randomly sheared fragments of *Drosophila* DNA from strain Canton-S¹², using as probe poly(A)⁺ RNA from late third-instar larval fat bodies^{1,3}. The restriction map shows the positions in 104 DNA of the endonuclease cleavage sites for *EcoRI* (labelled E), *SalI* (labelled S) and *BamHI* (labelled B)³. The relevant fragments produced by digestion of 104 DNA with both *EcoRI*



and *SalI* are labelled 1–5, and by digestion with *BamHI* are labelled 6–12. The location of the 5' terminus of *LSP-2* mRNA was determined by hybridizing fat body RNA with purified fragment 2 DNA, removing the single-stranded regions by digestion with *S*₁ nuclease⁶ and measuring the size of the remaining RNA–DNA hybrid by electrophoresis in agarose gel. Fragment 2 was purified after electrophoresis in 0.8% agarose gel of an *EcoRI/SalI* digest of 104 DNA, electroelution of the fragment into hydroxyapatite¹³ and recovery of the fragment from hydroxyapatite with 1 M sodium phosphate buffer pH 7; the purified fragment was dialysed at 4°C against 10 mM Tris/1 mM EDTA pH 7.6, precipitated with 95% ethanol and dissolved in 25 μl of a solution containing 0.4 M NaCl/0.04 M PIPES pH 6.4/1 mM EDTA/80% formamide⁶. The hybridization reaction was started by dissolving 100 μg total fat body RNA, heating the solution at 67°C for 10 min and then incubating at 52°C for 3 h to allow hybrids to form. The solution was then diluted into 10 vols of ice-cold *S*₁ nuclease reaction buffer and digested with 10 U of *S*₁ nuclease at 37°C for 35 min, as described by the manufacturer (BRL). The digestion was stopped and the hybrids precipitated with 95% ethanol in the presence of 10 μg tRNA. The precipitate was electrophoresed in a denaturing agarose gel, transferred to nitrocellulose paper, hybridized to ³²P-labelled 104 DNA (specific activity 10⁸ c.p.m. μg^{−1}) and autoradiographed, as previously described³. The resulting autoradiographs are shown beside the restriction map; lane a is a control containing an *EcoRI/SalI* digest of total 104 DNA, lane b contains the *S*₁ nuclease-resistant RNA–DNA hybrid formed with fragment 2. The exposure for a was 3 days, for b 10 days.

ing in size to the *LSP-2* mRNA was detected in fat body RNA but not in embryonic RNA, consistent with earlier results²; three other transcripts were detected in the embryonic RNA, of 2.6, 2.2 and 1.6 kb and designated *T1*, *T2* and *T3* respectively. The regions of 104 DNA homologous to each of the three transcripts were identified by using purified *EcoRI* fragments of 104 DNA as individual probes. Fragment 1, which contains the *LSP-2* coding region and 3.5 kb flanking the 5' end of the coding region, hybridized with the *LSP-2* transcript in fat body RNA and with the *T1* transcript in embryonic RNA; fragment 2 hybridized with transcripts *T1*, *T2* and *T3*; and fragment 3 with only *T3*. The doublet in the lane containing embryo RNA in Fig. 3b is probably the result of residual ribosomal RNA in the poly(A)⁺ RNA preparation. These bands appeared when the blot was hybridized to several probes unrelated to clone 104. Members of several other laboratories doing similar analyses have found similar, nonspecific bands at the ribosomal RNA position (personal communications).

To investigate whether *T1*, *T2* and *T3* were encoded by separate sequences or shared homology, three additional DNA fragments homologous to *EcoRI* fragment 2 were hybridized to the embryonic RNA blot of Fig. 3 (data not shown, see Fig. 3 legend). We observed that *EcoRI* fragments 2a and 2b were homologous to both *T1* and *T2*, whereas *T3* was homologous

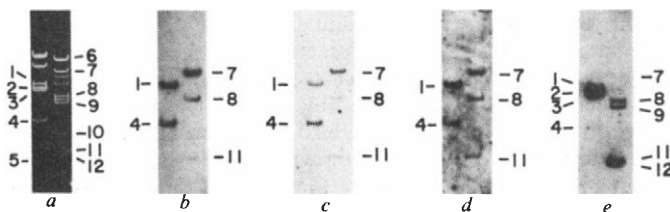


Fig. 2 Regions of 104 DNA that hybridize with RNA from oocytes, embryos and larval fat bodies. Samples containing 1.5 µg of 104 DNA were digested with either a combination of *EcoRI* and *Sal I* or *BamHI*, separated in 1% agarose gel and transferred to nitrocellulose paper³. The paper was prehybridized with 0.75 M sodium chloride/0.075 M sodium citrate/50% formamide/0.2% SDS/0.5 mg ml⁻¹ *Escherichia coli* rRNA at 41 °C for 2 h in sealed plastic containers, and the ³²P-labelled poly(A)⁺ RNA probe was added and the reaction continued for 48 h; the RNA was labelled with [γ-³²P]ATP and T4 polynucleotide kinase¹, yielding specific activities of 0.5–1.5 × 10⁷ c.p.m. µg⁻¹. The paper was washed by shaking for several hours at 41 °C in 0.75 M sodium chloride/0.075 M sodium citrate/50% formamide and then at room temperature in 0.75 M sodium chloride/0.075 M sodium citrate for 30 min. After a 1-min rinse in 0.3 M sodium chloride/0.03 M sodium citrate, the washed paper was wrapped in Saran plastic and autoradiographed for 3–7 days. *a*, Staining pattern with ethidium bromide; *b*–*e*, autoradiographic patterns obtained with poly(A)⁺ RNA probes from the following sources: *b*, stage 14 oocyte; *c*, gastrula; *d*, late (12 h) embryo; *e*, fat bodies from late third-instar larvae. In all the gels, the left lane contained the *EcoRI/SalI* digest and the right lane contained the *BamHI* digest of 104 DNA. The positions of the relevant bands are labelled as in Fig. 1.

to *EcoRI* 2c. As probe sequences separated by at least 4 kb are homologous to the 2.6-kb *T1* transcript, the DNA sequences encoding *T1* must be interrupted by at least one intervening sequence. The relationship between the *T1* and *T2* transcripts is uncertain, although the data are consistent with the coding of *T2* by a subset of the *T1* coding region. Additional tests for hybridization of the *LSP-2* and *T1* transcripts with restriction fragment *EcoRI* 1a from 104 DNA place the proximal terminus of the *T1* coding region 2.5–3.5 kb from the 5' end of the *LSP-2* coding region (data not shown).

Based on the proximity of sequences encoding *LSP-2* and *T1*, we determined whether or not *T1* transcription was augmented during *LSP-2* expression in the fat body. Samples containing poly(A)⁺ RNA from late embryos and fat bodies were electrophoresed in agarose gel, transferred to nitrocel-

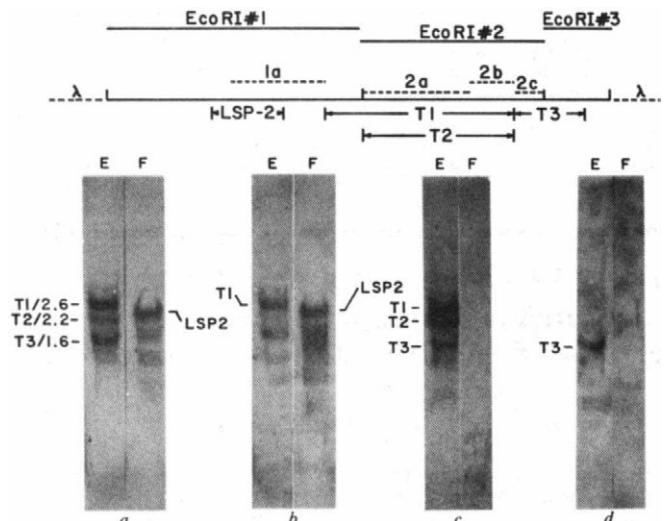


Fig. 3 Identification of transcripts from embryos and larval fat bodies that hybridize with 104 DNA. Samples containing either 23 µg poly(A)⁺ RNA from 12 h embryos or 10 µg total RNA from late third-instar larval fat bodies were glyoxalated, electrophoresed in an agarose gel and transferred to nitrocellulose paper^{3,14}. Hybridizations were done with *EcoRI* fragments of 104 DNA purified by the following procedure. An *EcoRI* digest of 104 DNA was electrophoresed in 1% agarose gel, using low-melting point agarose (BRL), and the DNA bands were detected by staining the gel with ethidium bromide; the bands containing the *EcoRI* fragments of *Drosophila* DNA were cut from the gel and heated at 65 °C for 20 min, and then 100 µl 10 mM Tris-HCl pH 7.6/1 mM EDTA containing 15 µg tRNA was added. The solution was brought to 0.2 M sodium acetate and shaken with an equal volume of phenol at 65 °C. The aqueous phase was shaken again with phenol at room temperature; the nucleic acid was precipitated from the aqueous phase in 95% ethanol. Each fragment was labelled by nick-translation, and the labelled probes were hybridized with the RNA transferred to nitrocellulose paper. The hybridizations were done successively using the same paper, with a regeneration step after each hybridization to remove the preceding probe¹⁴. Probes: *a*, total 104 DNA; *b*, *EcoRI* fragment 1; *c*, *EcoRI* fragment 2; *d*, *EcoRI* fragment 3. E lanes contain late embryo poly(A)⁺ RNA, F lanes contain total fat body RNA. The map positions of additional probes used in this analysis (1a, 2a, 2b, 2c) are indicated by the dotted lines above the *Drosophila* DNA insert in clone 104. The results from these hybridizations are explained in the text; the data are not presented. The map positions of the genes encoding *T1*, *T2* and *T3* are indicated by the lines below the *Drosophila* DNA insert.

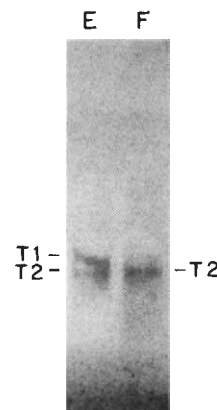


Fig. 4 Test for transcripts *T1* and *T2* in embryonic and fat body RNA. Samples containing 15 µg poly(A)⁺ RNA from late (12 h) embryos or 12 µg of poly(A)⁺ RNA from late third-instar larval fat bodies were denatured, electrophoresed in agarose gel and transferred to nitrocellulose paper^{3,14}. The *BamHI* fragment 8 (see Fig. 1) was purified as described for the *EcoRI/SalI* fragment 2 of Fig. 1, ³²P-labelled by nick-translation, hybridized to the RNA on nitrocellulose paper, and the paper was then autoradiographed for 3 days. Lane E contains embryonic RNA, lane F contains fat body RNA.

lulose and hybridized to ^{32}P -labelled *Bam*HI fragment 8 (Fig. 4). As predicted, this probe hybridized to the *T1* and *T2* transcripts present in the embryo RNA preparation, whereas only the *T2* transcript was detectable in the fat body sample. When the nitrocellulose filter was regenerated and hybridized to a probe containing the *LSP-2* coding region, the fat body RNA preparation hybridized ~90 times more strongly than the embryo RNA at the position of the *LSP-2* transcript (data not shown). Thus, despite a dramatic increase in the level of *LSP-2* transcript in late third-instar larval fat body tissue, the *T1* transcript does not accumulate.

There is no proof that the sequences encoding the *T1*, *T2* and *T3* transcripts are those flanking the *LSP-2* coding region. *T1*, *T2* and *T3* could be transcribed at other genomic sites, yet possess significant homology with clone 104 sequences. This is unlikely, however, because (see refs 2, 3, 7): (1) when genomic DNA was digested with several restriction endonucleases, electrophoresed in agarose gel and transferred to a nitrocellulose filter, after hybridization to ^{32}P -labelled clone 104 DNA the only bands detected on the filter were those predicted from the restriction endonuclease map of clone 104. (2) *In situ* hybridiz-

ations of tritium-labelled clone 104 DNA to salivary gland chromosomes have revealed only a single site of hybridization, the *LSP-2* locus. Thus, the *T1*, *T2* and *T3* transcripts are probably encoded by the region flanking the *LSP-2* gene. As *T1* transcripts are undetectable in third-instar larval fat body tissue, it seems that the factors which increase the expression of the *LSP-2* gene act locally, rather than regionally. This situation is very similar to the regulation of sequences encoding yeast histones⁸, *D. melanogaster* cuticle proteins⁹ and the enzyme dopa decarboxylase of *D. melanogaster*¹⁰. This is particularly interesting since the transcription of *LSP-2* seems to be augmented by the steroid hormone ecdysterone^{3,5}. DNA transformation studies have shown that sequences required for hormonal response can be closely linked to the coding regions they regulate¹¹. In the case of the *LSP-2* gene, the position of sequences encoding *T1* may delimit the regulatory regions.

B.Y. was supported by a NIH postdoctoral fellowship. M.L. was supported by grants to Alan Garen from the American Cancer Society and the National Institute of General Medical Sciences. This research was performed in the laboratory of Alan Garen.

Received 1 October 1981; accepted 22 March 1982.

1. Lepesant, J. A., Kejzarova-Lepesant, J. & Garen, A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5570–5574 (1978).
2. Levine, M., Garen, A., Lepesant, J. A. & Kejzarova-Lepesant, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2417–2421 (1981).
3. Lepesant, J. A. *et al.* *J. molec. appl. Genet.* (in the press).
4. Roberts, D. B., Wolfe, J. & Akam, M. E. *J. Insect Physiol.* **23**, 871–878 (1977).
5. Jowett, T. & Postlethwait, J. H. *Nature* **292**, 633–635 (1981).
6. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721–732 (1977).

7. Levine, M. thesis, Yale Univ. (1981).
8. Hereford, L. M., Osley, M. A., Ludwig, J. R. II & McLaughlin, C. S. *Cell* **24**, 367–370 (1981).
9. Snyder, M., Hirsh, J. & Davidson, N. *Cell* **25**, 165–178 (1981).
10. Hirsh, J. & Davidson, N. *Molec. cell. Biol.* **1**, 475–485 (1981).
11. Hynes, N. E., Kennedy, N., Rahmsdorf, U. & Groner, B. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2038–2042 (1981).
12. Maniatis, T. *et al.* *Cell* **15**, 687–701 (1978).
13. Tabak, H. F. & Flavell, R. A. *Nucleic Acids Res.* **5**, 2321–2332 (1978).
14. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).

Male and female mouse DNAs can be discriminated using retroviral probes

Sandra J. Phillips, Edward H. Birkenmeier, Robert Callahan* & Evan M. Eicher

The Jackson Laboratory, Bar Harbor, Maine 04609, USA

* Laboratory of Cellular and Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205, USA

Only two markers have been identified for the mouse Y chromosome: a gene that determines the male transplantation antigen (H-Y)¹ and sequences homologous to a satellite DNA originally isolated from a heterogametic female snake². To investigate the molecular organization of the murine Y chromosome and to understand its role in sex determination, more Y-specific sequences and genes must be identified. We have now compared restriction digests of DNAs from female and male mice using hybridization probes representing two different classes of murine retroviruses. Some viral DNA-containing fragments were present in male but not female DNAs. The male-specific fragments varied in copy number from 1 to as many as 100, and we estimate that in some *Mus* species, as much as 3% of the Y chromosome is composed of retrovirus-related sequences. To our knowledge this is the first report of (1) an association of retrovirus-related sequences with a mammalian Y-chromosome and (2) many copies of retrovirus-related sequences on the same chromosome.

DNAs were prepared from male and female mice representing six distantly related inbred strains of *Mus musculus* and three other *Mus* species³: *Mus poschiavinus*, *Mus spretus* and *Mus caroli*. *M. poschiavinus* and *M. spretus* were included because they interbreed with laboratory strains and *M. caroli* because its karyotype is virtually identical to that of *M. musculus*⁴.

*Eco*RI digests of mouse DNAs were blotted onto nitrocellulose⁵ and hybridized to the long terminal repeat (LTR) of a *M. musculus* amphotropic retrovirus⁶. The LTR sequence,

which shares homology with the AKR ecotropic virus LTR⁷, was isolated from unintegrated proviral DNA and subcloned in pBR322. This probe was chosen because it represents terminal viral DNA sequences and detects fusion fragments composed of viral and adjacent cellular sequences. Approximately 50 fragments in *M. musculus* DNA hybridize to the LTR (Fig. 1a, lanes 1–12), and they probably represent many of the endogenous type C retroviral genomes known to exist in *M. musculus*⁸. In the four inbred strains BALB/cWt, C57BL/10J, NZB/B1NJ and SM/J, a 5.5 kilobase (kb) *Eco*RI fragment was present in male but not female DNA (Fig. 1a, lanes 1–4, 7, 8, 11, 12). The relative intensity of this band suggested that it is present in one or a few copies. This dimorphic fragment was not observed in DNA from AKR/J and DBA/2J strains (Fig. 1a, lanes 5, 6, 9, 10).

Restriction fragments containing DNA sequences related to the LTR were weakly detected in the closely related species *M. spretus* (Fig. 1a, lanes 13, 14) and *M. poschiavinus* (Fig. 1a, lanes 15, 16). No male-specific bands were evident. DNA from the more distantly related species *M. caroli* did not hybridize with this probe, which is consistent with previous results^{9,10}.

A different hybridization pattern was observed when a sequence representing another class of endogenous xenotropic type C retrovirus was used to analyse these same DNAs (Fig. 1b). This class (designated type C-I) has been isolated from all species of *Mus* tested, except *M. musculus*, and can be distinguished from other laboratory strains of retroviruses both immunologically and on the basis of nucleic acid sequence homology^{10–12}. A C-I virus (designated M720) was isolated from *M. cookii* tissue culture cells and used to infect FEC cells. A clone of unintegrated circular proviral DNA was prepared. In *M. musculus* (Fig. 1b, lanes 1–12) and *M. poschiavinus* (Fig. 1b, lanes 15, 16), this clone of the entire *M. cookii* type C-I proviral genome hybridized to at least four fragments (14.5, 11.8, 7.5 and 4.3 kb) in male DNAs that were not present in female DNAs. Because the sum of the sizes of the male-specific bands is larger than a viral genome (8.5 kb), we conclude that these four bands are derived from more than one type C-I endogenous virus genome. The relative intensity of hybridization of these fragments to M720 suggests that each is present in multiple copies. Usually multiple-copy fragments represent

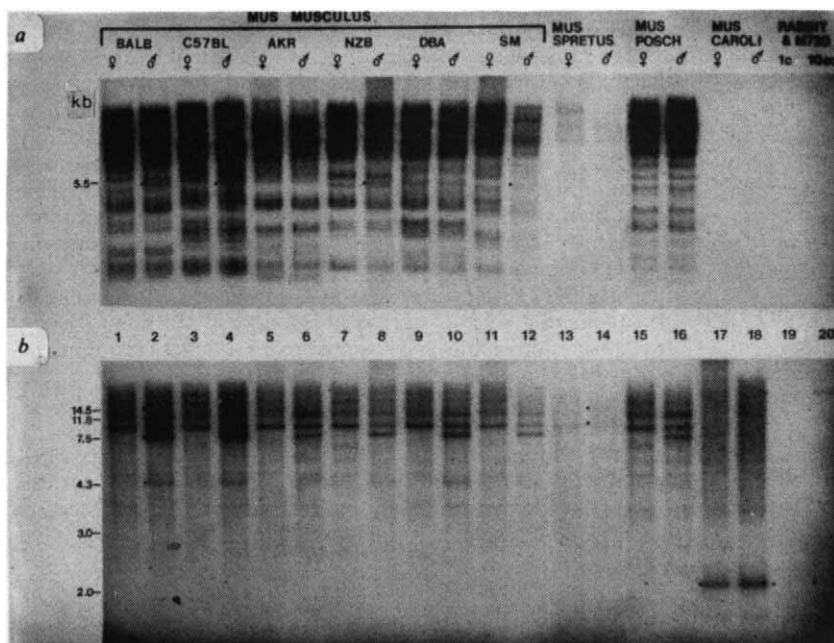


Fig. 1 Detection of sexually dimorphic retrovirus-related *Eco*RI restriction fragments in genomic mouse DNAs. Genomic DNAs were prepared from individual livers of inbred male and female siblings (lanes 1–12) or from closely related animals (lanes 13–18). DNAs (4 μ g) were digested with *Eco*RI, separated by electrophoresis in 1% agarose and transferred to nitrocellulose⁵. Probes were prepared by nick-translation to a specific activity of $1\text{--}3 \times 10^8$ d.p.m. 32 P per μ g DNA²⁵. Filters were incubated at 70 °C in the presence of $6 \times \text{SSC}$ for 16 h, and washed extensively at 52 °C in $0.1 \times \text{SSC}$, 0.05% SDS. Hybridizing fragments discussed in the text are denoted by ■ and the sizes of those fragments are given on the left of each panel. In *a*, mouse DNAs were probed with a pBR322 clone of the LTR of an amphotropic retrovirus from *M. musculus*⁶. Resolution was increased by exposing for 10 days to Kodak XAR film without intensifying screens. In *b*, the same DNAs were probed with a pBR322 clone of proviral DNA (M720) from a xenotropic retrovirus from *M. cookii*. Exposure was for 1 day at room temperature. Abbreviated strain and species names include BALB/cWt (BALB), C57BL/10J (C57BL), AKR/J (AKR), NZB/B1NJ (NZB), DBA/2J (DBA), SM/J (SM) and *M. poschiavinus* (MUS POSCH).

internal viral sequences, because flanking cellular restriction sites are presumed to be heterogeneous¹³. If all the *Eco*RI sites generating these multiple-copy fragments are internal viral restriction sites, then we are observing at least one unusually large (>14.5 kb) retrovirus genome. Alternatively, one or more conserved *Eco*RI sites flanking 'normal-sized' viral genomes could generate multiple-copy fragments larger than the viral genome itself.

The number of M720 virus-related sequences in male DNA from *M. musculus* was estimated by mixing cloned M720 proviral DNA with rabbit DNA at levels representing 1 or 10 copies of proviral DNA per cellular genome (Fig. 1b, lanes 19, 20). One copy of M720 proviral DNA per genome is barely detectable when hybridized with itself (lane 19). Ten copies of the viral genome are detectable, but the intensity is approximately one-tenth of that observed with the M720 male-specific fragments shown in Fig. 1b, lanes 1–12. From these and other results (data not shown) we conclude that there are ~100 copies of M720 related sequences in males from *M. musculus* inbred strains and *M. poschiavinus*.

Male-specific M720-related fragments were detected in *M. spretus* DNA (Fig. 1b, lanes 13, 14); however they hybridize less intensely with the probe than those identified in *M. musculus* male DNA (Fig. 1b, lanes 1–12). Whether this reflects a lower copy number or less homology is unresolved. Restricted cellular DNA from *M. caroli* also hybridized with M720 proviral DNA. Although *M. caroli* DNA does not appear to contain male-specific sequences unique to the M720 proviral genome, it does contain a 2.0 kb internal *Eco*RI fragment characteristic of the provirus (Fig. 1b, lanes 19, 20). We conclude that the M720-related sequences in the *M. caroli* genome are structurally more closely related to M720 than to related sequences in the other *Mus* species tested. In addition, the overall conserved pattern of the M720-related bands in *M. musculus* and *M. poschiavinus* DNAs is in striking contrast to the strain specificity observed with the LTR probe. These results are consistent with the previously noted conservation of the type C-I endogenous retrovirus-related sequences in cellular DNA from all species of *Mus*^{10,12}.

The presence of many copies of M720-related sequences on the Y chromosome of some *Mus* species can be explained in several ways. One hypothesis is that one or more independent viral integration events occurred in or near a specific sequence present many times on that chromosome. However, there is no evidence for site-specific integration of retroviruses¹⁴. Second, multiple copies of a viral genome could have resulted from successive intrachromosomal recombination events. Such a

model has been proposed to explain the presence of five copies of the avian leukosis retrovirus genome on a single chicken chromosome^{15,16}, and would explain the presence of 100 copies of M720-like virus sequences on the mouse Y chromosome. Third, we could be observing the result of amplification of a Y-chromosomal element containing virus-related sequences. The conserved pattern of M720-related dimorphic restriction fragments in *M. musculus* and *M. poschiavinus* suggests that either the integration or amplification event(s) responsible for multiple copies occurred recently in evolutionary terms, or these sequences have been conserved relative to other species. If conservation is responsible, these sequences may have a functional role. The role is not, however, sex determination, because mice that carry a *M. poschiavinus* Y chromosome (as well as the typical M720 male-specific fragments) can develop external female genitalia and ovaries¹⁷. It may be significant that males of several inbred strains have been shown to express a protein related to a type C virus envelope protein (gp70) in the epididymis and ductus deferens^{18,19}.

It remains to be demonstrated that the LTR and M720-related patterns of sexual dimorphism represent sequences from the Y chromosome. Although these sequences are present only when a Y chromosome is present, it is possible that the fragment dimorphism is a result of amplification of an autosomal or X-chromosomal sequence. Hybridization *in situ* will provide direct evidence.

We have shown that probes representing dispersed repetitive elements, such as retroviruses, are valuable tools for analysing large genomes. On Southern blots they can effectively reduce the complexity of the murine genome from 1×10^6 *Eco*RI fragments to ~50 fragments. Just as we have used retroviral probes to identify sequences from the Y chromosome, retroviral probes have been useful as molecular markers for identifying regions of the genome where deletions or insertions have occurred²⁰.

The Y chromosome is unique in the mammalian genome in that it is perpetually monosomic. With the exception of the region that pairs with the X chromosome during meiosis²¹, opportunities for recombination with other chromosomes are precluded. Markers such as these virus-related sequences offer a means to study the evolution of this chromosome, which may differ from that of the rest of the genome. Because the Y chromosome represents 2% of the haploid male genome (size = 1.5×10^9 base pairs)²⁴, the M720-related male-specific sequences would represent 3% of the chromosome [(100 copies $\times$ 8,500 base pairs / $0.02 \times 1.5 \times 10^9$) $\times$ 100]. The high concentration of virus-related DNA on the Y chromosome poses

many intriguing questions. Are these viral sequences present as a result of 'poor housekeeping' by the monosomic Y chromosome? Is the Y a 'target' chromosome for retrovirus integration or amplification? Are retroviral sequences involved in the function of this unusual chromosome?

We thank D. W. Bailey, N. A. Jenkins and L. L. Washburn for critically reading the manuscript. The Jackson Laboratory is fully accredited by the American Association for the Accreditation of Laboratory Animal Care. This work was supported by NIH grant GM 20919.

Received 28 January; accepted 15 March 1982.

- Eichwald, E. J. & Silmsker, C. R. *Transplant Bull.* **2**, 57-58 (1955).
- Jones, K. W. & Singh, L. *Hum. Genet.* **58**, 46-53 (1981).
- Marshall, J. T. & Sage, R. D. *Symp. zool. Soc. Lond.* **47**, 15-25 (1981).
- Miller, O. J. & Miller, D. A. in *Origins of Inbred Mice* (ed. Morse, H. C. III) 591-611 (Academic, New York, 1978).
- Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
- Rapp, U. R. & Todaro, G. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 624-628 (1980).
- Rands, E., Lowy, D. R., Lander, M. R. & Chattopadhyay, S. K. *Virology* **108**, 445-452 (1981).
- Jaenisch, R., Fan, H. & Croker, B. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4008-4012 (1975).
- Chattopadhyay, S. K., Lander, M. R., Rands, E. & Lowy, D. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5774-5778 (1980).
- Benveniste, R. E., Callahan, R., Sherr, C. J., Chapman, V. & Todaro, G. J. *J. Virol.* **21**, 849-862 (1977).
- Lieber, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **72**, 2315-2319 (1975).
- Callahan, R. & Todaro, G. J. in *Origins of Inbred Mice* (ed. Morse, H. C. III) 591-611 (Academic, New York, 1978).
- Steffen, D., Bird, S., Rowe, W. P. & Weinberg, R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4554-4558 (1979).
- Weinberg, R. A. *Rev. Biochem.* **49**, 197-226 (1980).
- Tereba, A., Crittenden, L. B. & Astrin, S. M. *J. Virol.* **39**, 282-289 (1981).
- Tereba, A. *J. Virol.* **40**, 920-926 (1981).
- Eicher, E. M., Washburn, L. L., Whitney, J. A. III & Morrow, K. E. *Science* (in the press).
- Del Villano, B. C. & Lerner, R. A. *Nature* **259**, 497-499 (1976).
- Lerner, R. A., Wilson, C. B., Del Villano, B. C., McConahey, P. J. & Dixon, F. J. *J. exp. Med.* **143**, 151-166 (1976).
- Jenkins, N. A., Copeland, N. G., Taylor, B. A. & Lee, B. K. *Nature* **239**, 370-374 (1981).
- Tres, L. L. *J. Cell Sci.* **25**, 1-15 (1977).
- Committee on Standardized Genetic Nomenclature for Mice *J. Hered.* **63**, 69-71 (1972).
- Disteche, C. M., Carrano, A. V., Ashworth, L. K., Burkhart-Shultz, K. & Latt, S. A. *Cytogenet. cell. Genet.* **29**, 189-197 (1981).
- Laird, C. *Chromosoma* **32**, 378-406 (1971).
- Maniatis, T., Jeffrey, A. & Kleid, D. G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1184-1188 (1975).

Messenger RNA of infectious pancreatic necrosis virus is polycistronic

P. P. C. Mertens* & P. Dobos

Department of Microbiology, University of Guelph, Guelph, Ontario, Canada N1G 2W1

Infectious pancreatic necrosis virus (IPNV) is an economically important fish pathogen. The virus particles contain two large segments of double-stranded (ds)RNA¹⁻⁴, which encode four different intracellular primary gene products^{5,6}. The smaller genome segment ('B') encodes the largest of these four proteins and the larger genome segment ('A') encodes the other three by a mechanism which does not seem to involve post-translational cleavage of a large precursor polypeptide *in vivo*⁷. Previous results^{8,9} have indicated that only two species of mRNA are produced, one from each genome segment⁸, and that one of these viral mRNAs (that produced from segment B; ref. 7) is monocistronic. The other mRNA, which is transcribed from segment A (ref. 7), encodes three proteins, which may be produced by a mechanism involving multiple initiation and termination sites for translation, rather than post-translational cleavage⁵⁻⁷. To distinguish between these two possible mechanisms, we have studied the translation *in vitro* of the denatured IPNV genomic dsRNA segments. We report here that the mRNA transcribed from genome segment A is polycistronic.

IPNV particles are icosahedral with a diameter of 60 nm. The dsRNA genome is bi-segmented (segments A and B of molecular weight (M_r) 2.5×10^6 and 2.3×10^6 , respectively¹⁻⁴; see Fig. 1A).

Three size classes of IPNV-specific polypeptides are synthesized in the same relative proportions throughout the infectious cycle, starting 3 h after infection⁵. These primary gene products are designated ICPs (infected cell proteins) and have molecular weights of 105,000 (105K), 62K, 31K and 29K (Fig. 1C^{5,7}). Of these four proteins, three subsequently undergo post-translational cleavage: (1) the 62K protein is cleaved via a 60K intermediate, to generate the major capsid protein VP54 (Fig. 1D); (2) part of the 31K internal virion protein is cleaved during maturation to generate a 29K polypeptide, and both

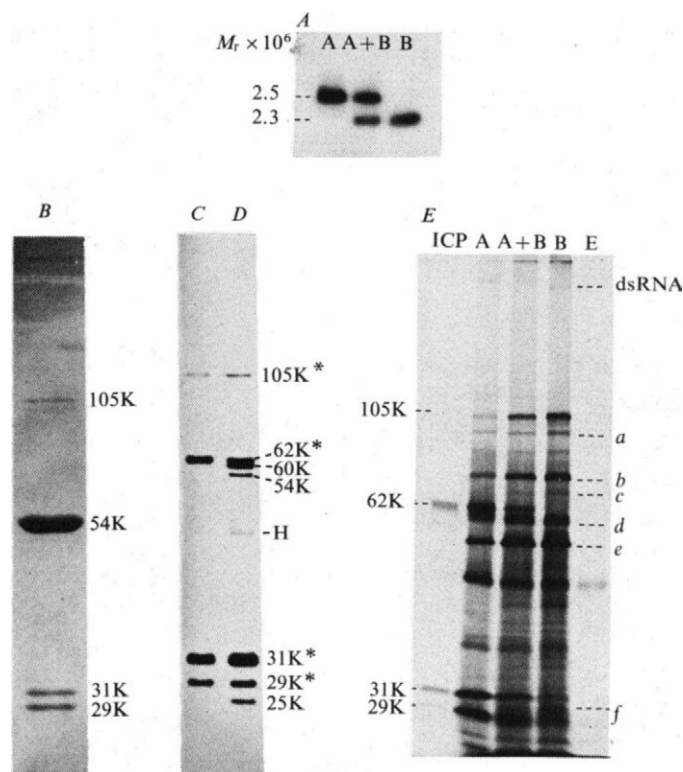


Fig. 1 Polyacrylamide gel electrophoresis of IPNV dsRNA, purified virus proteins, virus-specific proteins synthesized in infected cells, and protein products derived by translation *in vitro* of denatured IPNV dsRNA. Purified virus was digested with proteinase K and the genome segments were separated by electrophoresis in 10% polyacrylamide slab gels^{10,17}. The bands were detected by autoradiography (a small amount of ³²P-labelled dsRNA was added to the original sample), sliced out and electroeluted from the gel slices using methods published elsewhere¹⁰. A, a small amount of each fractionated genome segment was subjected to analytical gel electrophoresis followed by autoradiography to check its purity and homogeneity (note that due to trailing, segment A was contaminated with segment B). The molecular weights were computed from their electrophoretic mobilities relative to reovirus (Dearing 3) dsRNA run in parallel wells. B, proteins of purified IPNV analysed in 10% polyacrylamide slab gel followed by staining with Coomassie blue using procedures published elsewhere^{17,18}. C, D, autoradiograms of labelled virus-specific polypeptides analysed in 11% polyacrylamide slab gels. Infected CHSE cell monolayers were UV-irradiated 3 h after infection to reduce the cellular background and the cultures were labelled with ³⁵S-methionine 6 h after infection at 20 °C for 1 h (C) or 4 h (D) as described previously^{5,6}. The molecular weights of virus-specific proteins were computed from their electrophoretic mobilities relative to unlabelled standards run in parallel. H indicates residual host protein; the four primary gene products are marked with asterisks. E, cell-free protein synthesis was carried out in nuclease-treated rabbit reticulocyte lysates (NEN), primed with denatured dsRNA segments, in the presence of ³⁵S-methionine. 2 µg of dsRNA were precipitated in ethanol, washed with ether, dried and denatured in 2 µl of dimethyl sulphoxide at 55 °C for 15 min, then 100 µl of the complete reaction mixture was added and incubation was continued at 37 °C for 1 h. The products of translation were analysed using 11% polyacrylamide slab gels followed by autoradiography. ICP, labelled infected-cell proteins, synthesized in UV-irradiated CHSE cell cultures from 6 to 8 h after infection. A, *in vitro* translation products from dsRNA segment A; A+B, *in vitro* translation products from segments A and B; B, *in vitro* translation products from segment B; E, endogenous products of translation (no RNA added). The approximate molecular weights of the major products are indicated. a-f, The major products from incomplete translation of segment B. (Note the location of ³²P-labelled dsRNA near the top of the gel.)

* Present address: The Animal Virus Research Institute, Pirbright, Woking, Surrey GU24 0NF, UK.

precursor (VP31) and product (VP29) are found in the virion⁶ (Fig. 1C); (3) the only non-structural protein, 29K, is slowly degraded into small polypeptides via a 25K intermediate⁶ (Fig. 1D). The 105K protein is synthesized in small quantities and is incorporated into virions (VP105) without apparent changes in molecular weight (Fig. 1B-D). Peptide mapping has shown that the four primary gene products have different amino acid sequences⁶.

Using hybrid viruses of two parental serotypes, it has been shown that genome segment B encodes ICP 105K and A encodes the other three primary gene products⁷. The question arises as to how three unrelated proteins are produced from a single large genome segment. We considered the following possible mechanisms: (1) post-translational cleavage of a large, precursor polypeptide; (2) transcription of subgenomic mRNAs from genome segment A; (3) transcription of genome-length mRNA containing multiple initiation and termination sites for protein synthesis.

Several studies have provided evidence against mechanisms 1 and 2^{5,6,8,9}. For example, we have been unable to detect

polyprotein precursor(s) in IPNV-infected cells by pulse-chase experiments or by using amino acid analogues, protease inhibitors, ZnCl₂ or supra-optimal temperatures⁵ (experiments with pactamycin also gave negative results; P.D., unpublished results). In addition, the relative molar ratios of the primary gene products indicated that the frequency of their synthesis was inversely related to their molecular weights, suggesting that they are the products of different cistrons⁵. Furthermore, during intracellular replication of IPNV, only two species of genome-length 24S viral mRNAs were synthesized, which hybridized to the two genome segments⁸. The virion-associated polymerase was shown to synthesize *in vitro* two species of single-stranded (ss)RNA which were the same and which co-migrated during electrophoresis with the species of viral message synthesized *in vivo*⁹. Thus no discrete subgenomic mRNAs were detected either *in vivo*⁸ or *in vitro*⁹.

To investigate the third possible mechanism (that is, multiple initiation and termination sites on a single mRNA species), we used an *in vitro* protein synthesizing system. Denatured virion dsRNA was used to provide translation templates¹⁰ in rabbit

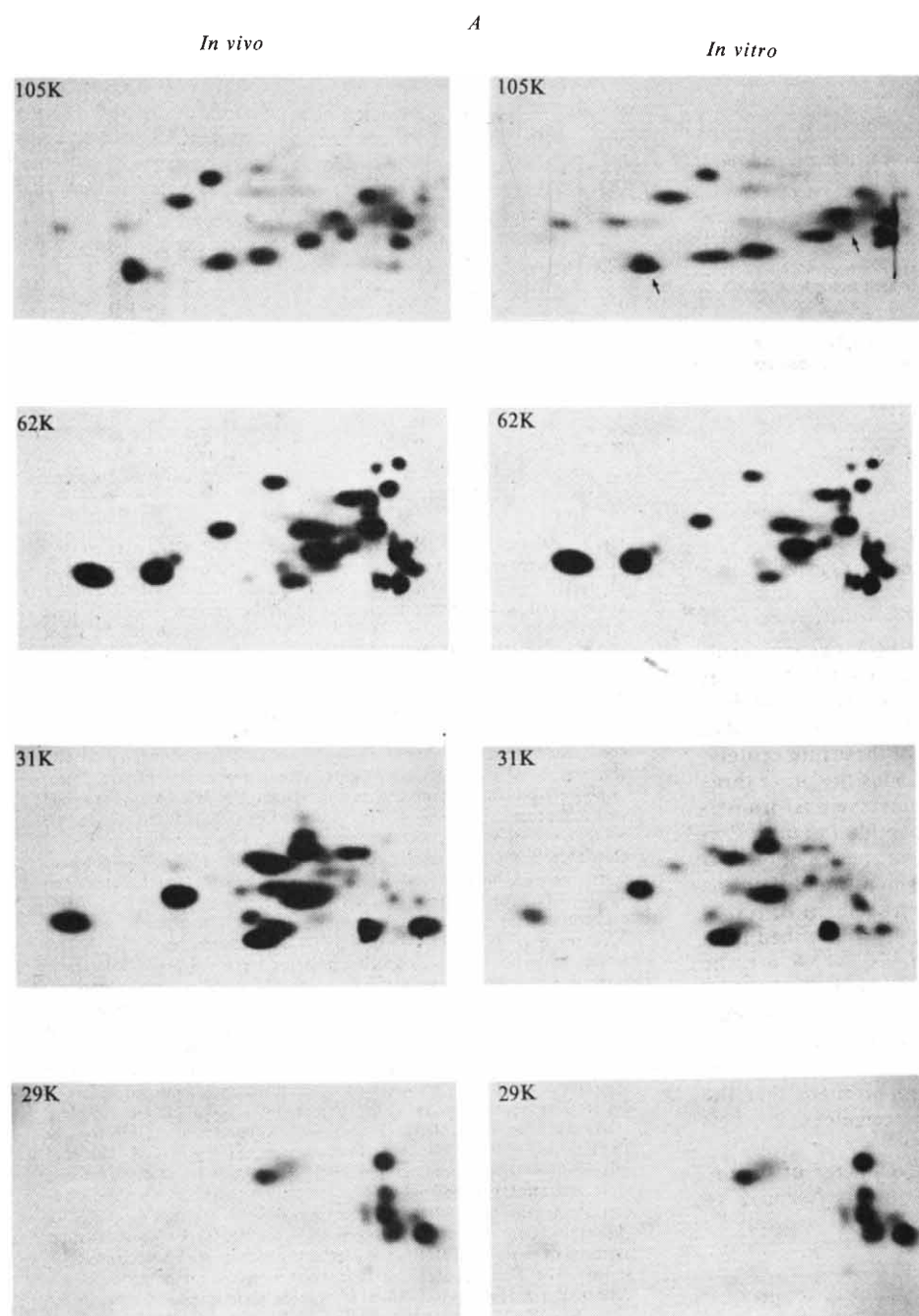


Fig. 2 Tryptic peptide mapping of ³⁵S-Met-labelled polypeptides produced *in vivo* and *in vitro*. **A**, polypeptide products identified as bands 105K, 62K, 31K and 29K on the autoradiograph shown in Fig. 1E were excised from preparative polyacrylamide slab gels, and subjected to trypsin digestion. The labelled peptides were separated in two dimensions by thin-layer electrophoresis and chromatography, then visualized by autoradiography as described previously^{6,19}. Samples were applied to the bottom right-hand (anode) corners. Arrows indicate peptides of *in vitro*-produced 105K protein that have different electrophoretic mobilities compared with their *in vivo*-synthesized counterparts. **B**, tracings of two-dimensional tryptic peptide maps for mixed digests of polypeptides 62K+31K (top) and 31K+29K (bottom). Protein processing, digestion, thin-layer electrophoresis, chromatography and autoradiography were performed as described for **A**. Open regions represent peptide spots of 62K and 29K proteins; cross-hatched areas are tracings of peptide spots of 31K polypeptide; and solid areas indicate spot overlaps (electrophoresis upper panel, 3.5 h; lower panel, 5 h).

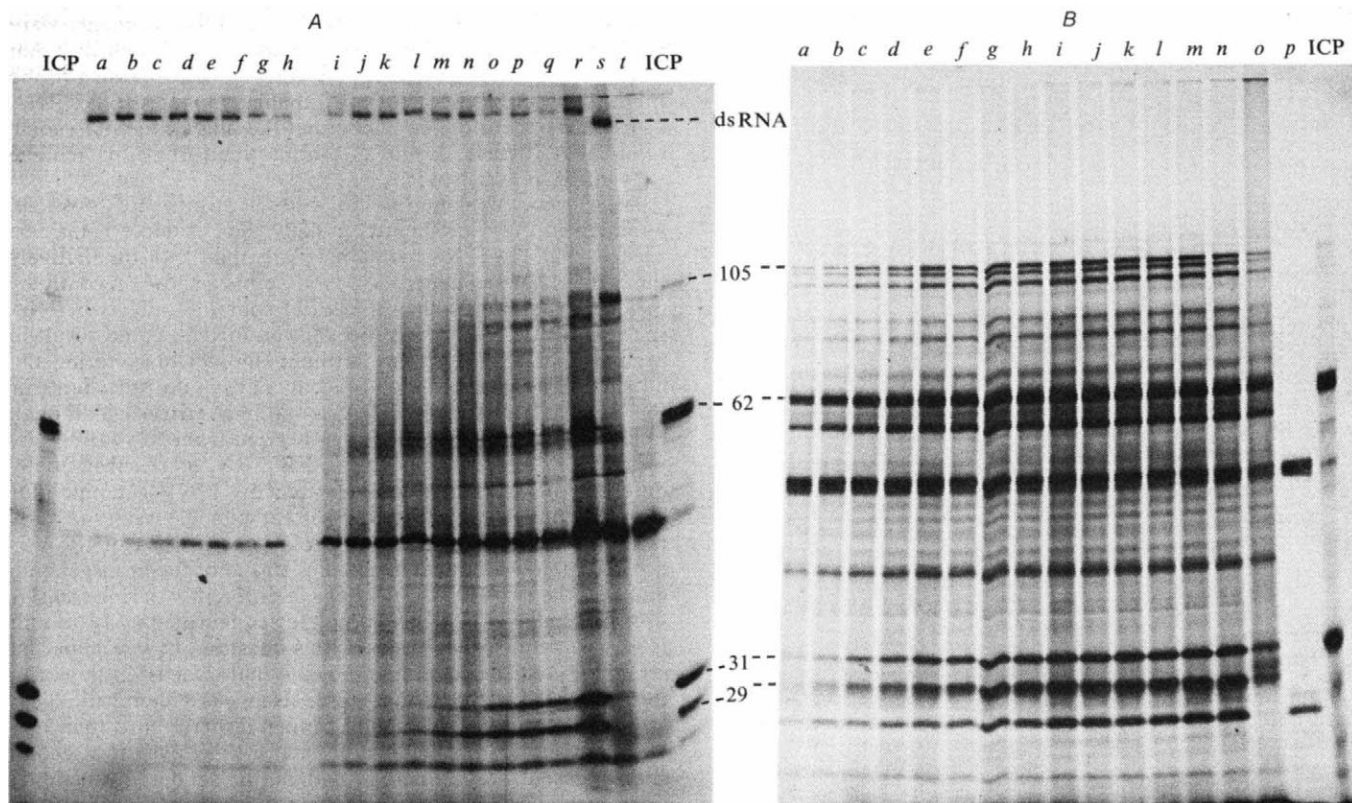


Fig. 3 Analysis of proteins synthesized *in vitro* by translation of denatured IPNV dsRNA in the presence of ^{35}S -methionine. **A**, time course of protein synthesis; cell-free protein synthesis was initiated as described for Fig. 1E. At intervals, 2- μl samples were removed from the reaction mixture and denatured by boiling in SDS-containing electrophoresis sample buffer. The labelled products were analysed in 11% polyacrylamide slab gels as described for Fig. 1E. *In vitro* translation primed with dsRNA segment A was stopped after the following incubation periods (min): a, 1; b, 2; c, 3; d, 4; e, 5; f, 6; g, 7; h, 8; i, 10; j, 12; k, 14; l, 16; m, 18; n, 20; o, 25; p, 30; q, 45; and r, 60. s, *In vitro* translation products from genome segment B, 60 min. t, Endogenous product (no RNA added), 60 min. ICP, labelled infected cell proteins as in Fig. 1E. (Note the position of ^{32}P -labelled dsRNA segments near the top of the gel.) The approximate molecular weights of the primary gene products are indicated. **B**, pulse-chase experiment; cell-free protein synthesis was initiated as described for Fig. 1E using both segments of dsRNA. At various intervals, 5 μl were removed from the reaction mixture, a 1,000-fold excess of unlabelled methionine was added and incubation continued for a total of 60 min. The labelled products were analysed by electrophoresis as described for Fig. 1E. Excess unlabelled methionine was added to samples of the reaction mixture at the following times (min): a, 2; b, 4; c, 6; d, 8; e, 10; f, 12; g, 14; h, 16; i, 18; j, 20; k, 30; l, 40; m, 50; n, 60; and o, 60, followed by incubation at 37 °C for 16 h. p, No RNA was added (endogenous, 60 min). ICP, labelled infected cell proteins as in Fig. 1E.

reticulocyte lysate. The two dsRNA genome segments were separated by preparative polyacrylamide gel electrophoresis (Fig. 1A) and recovered by electroelution¹⁰. Due to trailing, segment A was contaminated with B even after two cycles of electrophoresis, whereas segment B was homogeneous. The results in Fig. 1E show that proteins having similar electrophoretic mobilities to the four primary gene products produced in infected cells were also synthesized *in vitro*. Translation of genome segment B produced polypeptide 105K and several smaller polypeptides (labelled a-f) which were the result of premature termination of translation, as determined by peptide map comparisons (data not shown). The 105K protein produced *in vitro* exhibited slightly faster electrophoretic mobility than the same protein made in infected cells. This may be due to the addition *in vivo* of prosthetic groups, for example by glycosylation, methylation or phosphorylation, which may have reduced the electrophoretic mobility of ICP 105K. This electrophoretic difference was also reflected in minor variations in the otherwise identical peptide patterns of the *in vivo* and *in vitro* synthesized polypeptide 105K (Fig. 2A, arrows). This polypeptide is a minor product *in vivo*^{5,6} (Fig. 1C), but was produced in large quantities *in vitro* suggesting that not all of the intracellular translational control mechanisms are operative.

Translation of genome segment A resulted in three additional polypeptides having identical electrophoretic mobilities to ICP 62K, 31K and 29K (Fig. 1E). Due to contamination of the preparation with the very efficient template B RNA, polypeptide 105K and its incomplete translation products were also

found, although in reduced amounts. The tryptic peptide maps of the *in vitro* synthesized proteins 62K, 31K and 29K are identical to those of the equivalent proteins produced in IPNV-infected cells (Fig. 2A).

Although rapid protein processing could not be detected *in vivo*⁵, the high degree of radiolabelling, and the synchronous initiation of translation achieved *in vitro* prompted us to re-examine the possibility of post-translational cleavage. Analysis of the time course of protein synthesis where *in vitro* translation primed with dsRNA segment A was terminated at intervals between 0 and 60 min of incubation, showed that the order of appearance of polypeptides was 29K, 31K, 62K (Fig. 3A). The time for completion of polypeptide chains was proportional to their length, that is, smaller proteins were completed sooner (10 min) than longer ones (16 min). A large molecular weight precursor polypeptide could not be detected. Such a pattern is compatible with three independent initiation sites; it could also arise, however, if the order of polypeptides in a putative precursor were NH_2 -29K-31K-62K-COOH and if rapid proteolytic cleavage occurred as soon as the ribosome had passed the 3' end of a given cistron.

To distinguish between these two alternative mechanisms, we performed the following *in vitro* pulse-chase experiment. Translation was initiated in the presence of ^{35}S -methionine, then at 2-min intervals a 1,000-fold excess of unlabelled methionine was added and incubation continued for 1 h to complete chain elongation. The data in Fig. 3B show that all three proteins became labelled even during the shortest pulse

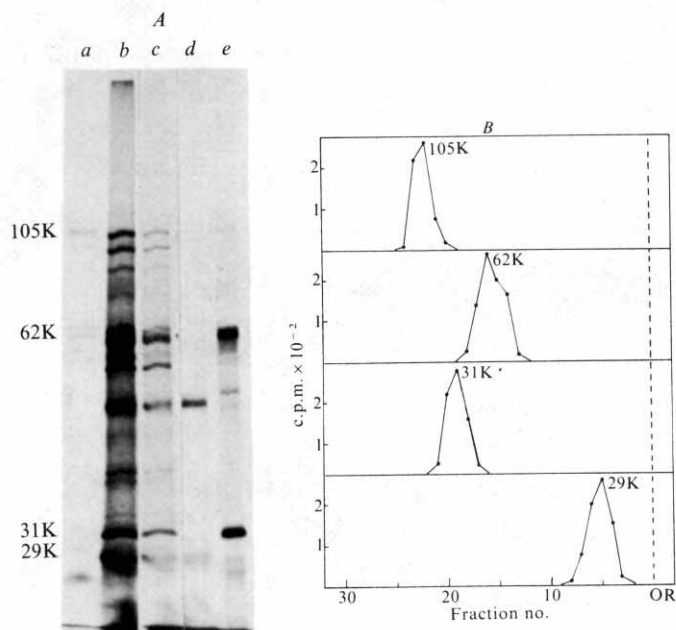


Fig. 4 Analysis of proteins synthesized *in vitro* by translation of denatured IPNV dsRNA in the presence of ^{35}S -fMet-tRNA^{Met}. **A**, cell-free protein synthesis was carried out using both segments of denatured dsRNA in the presence of ^{35}S -fMet-tRNA^{Met} (as described for Fig. 1E). The labelled products were analysed in 11% polyacrylamide slab gels and visualized by fluorography. **a**, ^{35}S -Met-labelled *in vitro* translation product; **b**, ^{35}S -fMet-labelled *in vitro* translation product; **c**, same as **b** but exposed for a shorter period; **d**, no added RNA (endogenous product); **e**, ICP as in Fig. 1E. The molecular weights of the primary gene products are indicated along side. **B**, Thin-layer electrophoresis of the ^{35}S -fMet-labelled tryptic peptides of polypeptides 105K, 62K, 31K and 29K. Processing and tryptic digestion of fMet-labelled polypeptides were performed as described in Fig. 2 legend. The digests were applied near the anodic end of a thin-layer silica gel plate and analysed in parallel by electrophoresis in pyridine/acetic acid/water (1:10:100 v/v/v) pH 3.5 at 450 V for 6.5 h. The dried gel was sliced into 5-mm fractions and each fraction was placed in a toluene-based scintillation cocktail and counted in a liquid scintillation counter. The radioactivity plotted represents net counts greater than twice the background. The broken line marks the origin (OR), where samples were applied. Electrophoresis was from right to left, with the anode on the right.

(2 min). These results are consistent with simultaneous initiation of translation at three independent sites (one for each protein). If initiation had occurred at a single site, followed by cleavage of the nascent chain, then the proteins should have been labelled sequentially from the N-terminus of the putative precursor molecule, with increasing pulse time.

It may be argued that the data are also compatible with the three proteins having common N-termini (overlapping genes), in which case the gene encoding 29K would be situated within the gene for polypeptide 31K, which in turn would be located within the gene for 62K protein. If this were the case, the peptide maps of 29K and 31K proteins would be virtually identical and would show a 50% overlap with that of 62K protein; however, this is not the case as shown in Fig. 2B (see also Fig. 2A).

When the four proteins were synthesized *in vitro* in the presence of ^{35}S -fMet-tRNA^{Met}, the N termini of all four polypeptides became labelled with ^{35}S -fMet, indicating independent initiation of translation (Fig. 4A). Furthermore, the labelled (N-terminal) tryptic peptides of these proteins exhibited different electrophoretic mobilities when analysed by thin-layer gel electrophoresis (Fig. 4B). These results do not, however, exclude the possibility of the three genes initiating and overlapping in the different reading frames on the mRNA.

That intact renatured ^{32}P -labelled dsRNA segments were detected after translation *in vitro* (Figs 1E, 3A), indicates that there was neither detectable breakdown nor processing of the input denatured dsRNA into smaller monocistronic RNA species. We therefore conclude that the plus strand of genome segment A (that is, the mRNA transcribed from this genome segment) is polycistronic.

Alternative explanations of these results are possible, although in our view highly unlikely. For example it may be argued that (1) IPNV contains more than two qualitatively different RNA segments; however we have shown recently by genetic experiments that the virus contains only two RNA species⁷. (2) One or two splices in the mRNA could result in a change of reading frame without appreciably altering the molecular weight of the RNA, but there is no precedent for this in the case of dsRNA-containing viruses. Although influenza virus NS₂ mRNA is synthesized from NS₁ mRNA by processing¹¹, the sizes of the two NS mRNAs are substantially different. It also seems unlikely that the IPNV genomic plus strand would be spliced correctly in reticulocyte lysates *in vitro*. (3) It may also be argued that the *in vivo* mRNA of IPNV is not completely identical to the plus strand of the virus genome and that the two RNAs may behave differently during translation. We believe that only complete sequence analysis of both RNAs can conclusively answer this question, however, partial sequence studies of reovirus indicate that, at least in this case, the viral mRNA and genomic plus strands are identical¹².

Eukaryotic mRNAs usually cannot display more than one initiation site for protein synthesis at a time, even if a second site is present on the mRNA and can be rendered functional by a change in configuration, as in the case of togaviruses¹³. The mRNA of Kunjin virus (a member of the flaviviruses) may contain multiple initiation sites for translation; however, this has not been demonstrated by *in vitro* translation using ^{35}S -fMet labelling, nor has it been confirmed by other investigators¹⁴. Recently, indirect evidence has indicated that the late 16S mRNA of simian virus 40 is polycistronic, encoding the agnoprotein and VP1 (ref. 15).

Although the results given here indicate that translation of the three proteins encoded by genome segment A of IPNV is initiated independently, we have no evidence concerning the relative position of the different cistrons on the mRNA. It is possible that the three coding regions are located in different reading frames and therefore overlap to some extent. Work is in progress to determine the order of cistrons on the A segment of the IPNV genome. We have also begun *in vitro* translation experiments to elucidate the mechanism of gene expression of two other bi-segmented dsRNA-containing viruses, that is, *Drosophila* X virus and infectious bursal virus of chickens¹⁶.

This work was supported by the Natural Sciences and Engineering Research Council of Canada. We thank Drs F. Brown, C. Payne, D. Rowlands, D. Sangar and R. Macdonald and Mr B. Underwood for helpful criticism during the preparation of this manuscript.

Received 17 November 1981; accepted 22 March 1982.

- Cohen, J., Poinard, A. & Scherrer, T. *J. gen. Virol.* **21**, 485-498 (1973).
- Dobos, P. *Nucleic Acids Res.* **3**, 1903-1924 (1976).
- Macdonald, R. D. & Yamamoto, T. *J. gen. Virol.* **34**, 236-247 (1977).
- Dobos, P., Hallett, R., Kells, D. T., Sorensen, O. & Rowe, D. *J. Virol.* **22**, 150-159 (1977).
- Dobos, P. *J. Virol.* **21**, 242-258 (1977).
- Dobos, P. & Rowe, D. *J. Virol.* **24**, 805-820 (1977).
- Macdonald, R. D. & Dobos, P. *Virology* **114**, 414-422 (1981).
- Somogyi, P. & Dobos, P. *J. Virol.* **33**, 129-139 (1980).
- Mertens, P. P. C., Jamieson, P. J. & Dobos, P. *J. gen. Virol.* **57**, 47-56 (1982).
- McCrae, M. A. & Joklik, W. K. *Virology* **89**, 578-593 (1978).
- Inglis, C. S., Gething, M. J. & Brown, C. M. *Nucleic Acids Res.* **8**, 3575-3589 (1980).
- McCrae, M. A. *J. gen. Virol.* **55**, 393-403 (1981).
- Kennedy, S. I. T. *J. molec. Biol.* **108**, 491-551 (1976).
- Westaway, E. G. *Virology* **80**, 320-335 (1977).
- Gilbert, J., Normura, S., Anderson, C. W. & Khoury, G. *Nature* **291**, 346-349 (1981).
- Dobos, P. *et al. J. Virol.* **32**, 593-605 (1979).
- Laemmli, U. K. *Nature* **227**, 680-685 (1970).
- Carstens, E. B. & Weber, J. *J. gen. Virol.* **37**, 453-474 (1977).
- Elder, J. A., Pickett, R. A. II, Hampton, J. & Lerner, R. A. *J. Biol. Chem.* **252**, 6510-6515 (1977).

BOOK REVIEWS

The games mathematicians play

C.W. Kilmister

THE business of games is too serious to be left to the players, and this is the main message of *Winning Ways*. It is in the long tradition of Rouse Ball, Dudeney and Lucas, but a great deal further along the line. Some years ago one of the authors, John Conway, published *On Numbers and Games* (Academic, 1976). The first half of that book contained a highly idiosyncratic foundation of the number-system, generating the integers, rationals and reals by a single process which yields them in other than the usual order. There was also an extension of the system to other than "usual" numbers. The second half applied this almost as an afterthought to certain games; but Conway claimed that this second half was really the main part of the book. *Winning Ways* generously redresses the balance. It is hard to characterize the content, except to say that it is mathematical games, understood as broadly as possible, and with a good deal of mathematical lessons deduced from them.

The first part of Vol. 1 discusses the types of game described in Conway's earlier book. In these games two players move alternately (often from a particular starting position) according to rules allowing options, until a player loses by being unable to move. There is complete information — both players know what is happening. So this is precisely *not* the "Theory of Games", and no intervention of chance is allowed. Nim is the prototype of such games. (Elsewhere in the book some of these conditions are relaxed, but never the restriction to complete information and no chance.)

There are two foci of interest: the technique of attaching numbers to positions in games to assess their values, and the analysis of the structure of new games formed from the sums of known and analysed ones. The first of these is so simple as to excite (initial) disbelief. If in any position P the two players have options leading to positions of values, $a, b, c, \dots$, or $d, e, f, \dots$ respectively, the value of P is written $\{a, b, c, \dots \mid d, e, f, \dots\}$. This value may or may not be a number. If $a, b, c, \dots, d, e, f, \dots$ are all numbers, the value of P is the simplest number (to find out exactly what simplest means you have to buy the book), if it exists, strictly greater than $a, b, c, \dots$ and strictly less than $d, e, f, \dots$. The beauty of this system is that it needs no start; for $\{ \mid \}$, in which neither player has a move, is a zero position i.e.

Winning Ways, for Your Mathematical Plays. By Elwyn R. Berlekamp, John H. Conway and Richard K. Guy. Two vols, pp.850. Vol. 1 *Games in General*, hbk ISBN 0-12-091150-7, pbk ISBN 0-12-091101-9; Vol. 2 *Games in Particular*, hbk ISBN 0-12-091152-3, pbk ISBN 0-12-091102-7. (Academic: 1982.) Each volume hbk £31.40, \$64.50; pbk £10.80, \$22.50.

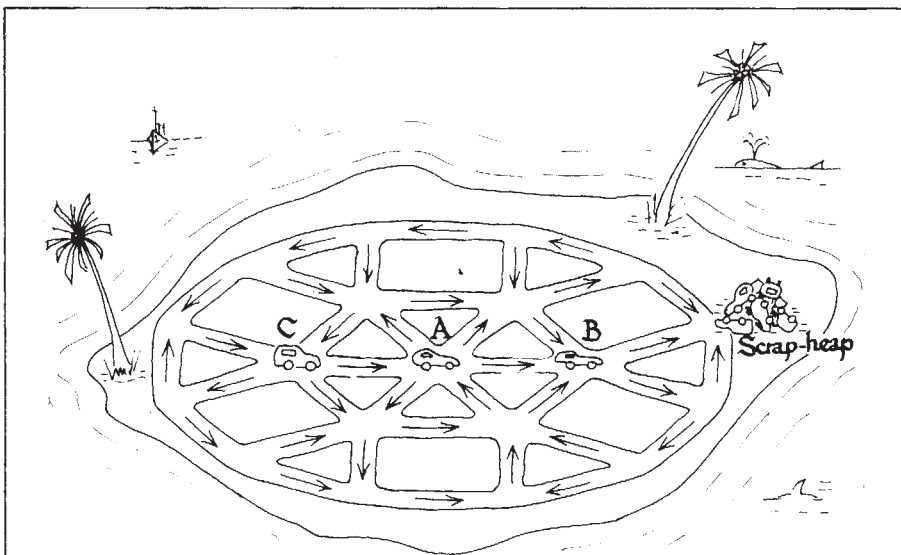
$\{ \mid \} = 0$. Then $\{0 \mid \} = 1$, $\{1 \mid \} = 2$, $\dots$, $\{0 \mid 1\} = \frac{1}{2}$. When the value of P is a number, it has an immediate interpretation as the number of moves advantage. (Then $\frac{1}{2}$ would correspond to a one move advantage in two copies of the game played together — and this leads on to the idea of the sum of two games, in which they are played side by side, and each player, when he moves, may choose to move in either game.) But even if the value is not a number, and so belongs to a natural extension of the rational numbers including infinitesimals, the ordering can be extended to say something useful.

The second part of the book, which concludes Vol. 1, goes on to games where some conditions are relaxed: you can move in several components of a compound game, or must move in all, or there are infinitely many positions, or the game may not stop. The main part of Vol. 2 applies these ideas to games of various kinds with coins, pencil-and-paper or on a board.

The approach is typified by noughts-and-crosses, of which a concise yet complete analysis is given. As the authors say:

We have no doubt that most of our readers were bright enough as children to discover that [a tied game] always happens when the game is properly played, and only the authors of books like *Winning Ways* retain sufficient interest to study the game in any detail.

But they go on, with approval, to quote Martin Gardner's implied castigation of "many players [who] have the mistaken impression that because they are unbeatable they have nothing more to learn". These two views summarize the attitudes and lessons of the book, applied to many



An example of the games — and humour — in *Winning Ways*, taken from Vol. 1. Above is a view of Corral Island, which was a Charming Antipodean Beauty Spot before the Corral Automotive Betterment Scheme, financed by the island's three wealthy landowners, built those expensive new super-highways. The island's two political parties (Left and Right) are now running for election and the gallop polls predict a walk-over for whichever party puts the last car on the scrap-heap. The parties have already enforced the one-way system indicated and each will alternately bribe a chauffeur to drive his car along one highway (in the proper direction) to the next intersection. You will see that there is no legal move away from the scrap-heap. How should we advise the ruling party, which must make the first move?

more complex games.

Finally, in Vol. 2, are two chapters on one-person games, and a final one on a no-person game. This is Conway's famous "Game of Life" about which a lot of recent news is here available: for example, that every sufficiently well-posed mathematical question can be converted to a question about the ultimate destinies of some Life configurations, so that Life is as insoluble as mathematics. Again, to quote: "Many computers have been programmed to play the game of Life. We shall now return the compliment by showing how to define Life patterns that can imitate computers". One of the one-person game chapters is solitaire and its extensions, the second on puzzles (the Tower of Hanoi, eight concise pages on Rubik's cube and so on).

The books are elegantly printed and written in a light-hearted vein, with a high density of puns per page, some outrageous but all appropriate. There is a wealth of illustrations — many of them are in colour — for which particular praise is deserved. A good index begins with an essential glossary of symbols.

If you make a practice of playing mathematical games this book is a must; you dare not play again without it. But even if you are not among those who indulge you ought to read it, as evidence of the insights into mathematics and physics you're missing. □

C. W. Kilmister, a Professor in the Department of Mathematics at King's College, University of London, is totally incompetent at any of the games mentioned. He has, however, used the material in Conway's earlier book to try to provide a combinatorial reformulation of quantum mechanics.

Uncovering discovery

I. Bernard Cohen

The Social Basis of Scientific Discoveries. By Augustine Brannigan. Pp.212. Hbk ISBN 0-521-23695-9; pbk ISBN 0-521-28163-6. (Cambridge University Press: 1982.) Hbk £12.50, \$24.95; pbk £4.95, \$9.50.

MOST of this new book by Augustine Brannigan is devoted to a review of theories about the nature of discoveries in science. The author begins with a description of various psychological theories of scientific discovery, limiting himself to the work of philosophers of science and not really investigating the accounts of psychologists. He then switches to the "emergence of a social model of discovery", which leads him into the fundamental question of heredity versus environment, nature versus nurture.

For convenience, Brannigan begins with Francis Galton's *Hereditary Genius*,

intended to show the way in which extraordinary talents ran in families. In antithesis to this, he presents the work of the anthropologist A. L. Kroeber, who stressed the importance of social environment, noting how great periods of history "emerged as a function of cultural development". In particular, Kroeber emphasized the significance of multiple discoveries in science as proof of the importance of the general intellectual or cultural milieu as opposed to that of individual acts of genius. Recently the argument about multiple discovery has been advanced and developed by Robert K. Merton and his students and co-workers. Brannigan further develops Merton's ideas in a most interesting and incisive manner and then contributes a thought-provoking analysis of the work of Mendel in relation to its alleged multiple rediscovery.

This leads Brannigan into an account of the nature of discovery itself and of what he calls "folk reasoning" concerning the theory of scientific discovery. There is a concluding discussion of the sociological features of discovery, where the author develops his own views of the subject and includes a model in which "the normative considerations described by Merton" are enlarged by consideration of the "cognitive grounds" for disputes associated with multiple discoveries. The author can only be applauded for this attempt to add a new dimension to the current debate, but it is greatly to be regretted that this important idea is not developed by fully worked-out examples. And it is the same for the author's programme for analysing "the rhetorical and organizational status of methods in science". It is very frustrating that in the final page or so he no more than hints at research concerning "aspects of the scientific community, its productivity, its stratification system, and factors which influence these things". He introduces a number of suggestions, but they are shot forth with rapid fire, not illustrated or developed.

Indeed, one of the great criticisms to be levelled against this book — and against so many works of this type — is that there are passing references to many individuals and episodes in the development of science, but nothing more. Only in the case of Mendel does Brannigan attempt to do any serious history. He does, to be sure, introduce some aspects of the discovery of oxygen, but even here there is no attempt to go deeply into either the primary sources or the major scholarly investigations which could have further illuminated this topic, both for himself and the reader. As a result, the book will appeal more to those interested in theoretical questions of the sociology or philosophy of science, than to scientists or those whose primary concern is science itself. □

I. Bernard Cohen is Victor S. Thomas Professor of the History of Science at Harvard University.

Too reproductive

Colin Stern

Reproductive Immunology. Edited by Norbert Gleicher. Pp.492. ISBN 0-8451-0070-X. (Alan R. Liss/Heyden: 1981.) \$86, £68.

THIS curious volume was cobbled together from some of the papers given at the first S.B. Gusberg Seminar on Reproductive Immunology, in June 1980 in New York, and from contributions solicited after the event which give it a strange chiaroscuro feel. The editor seems to have aimed at obstetricians and perinatologists, as well as immunologists and developmental biologists. There are several very good contributions but also some appalling ones. The book begins with three introductory chapters: two, on immunogenetics and immunoassay are good; the third, "Introduction to Immunology", is dreadful. It opens with a reasonable historical introduction, but the rest is full of errors of fact and misconceptions too numerous to detail. One is left asking whether this was necessary and why the author, an obstetrician whose knowledge of immunology is patently sketchy, was asked to write it.

There follow an excellent discussion of V-gene activation, solid contributions on immunoglobulin synthesis and lymphoid cell ontogeny, and a poor chapter on the development of the fetal immune system: how could the authors fail to discuss the acquisition of self-tolerance? The next part, "Pregnancy Immunology", begins with a brilliant overview of the immunobiology of the materno-fetal relationship by Rupert Billingham: clear and well-written, it makes much of the rest of the book seem redundant. There are other informative chapters here, however, on syncytiotrophoblast, uterine lymph node reactivity, maternal lymphocyte reactivity and breast milk. The next section starts well — Bill Pollack's chapter on Rhesus haemolytic disease — which makes further comment (from Cherry) unnecessary.

As is typical of the unevenness of the book as a whole, the following section contains two good contributions, on SLE in pregnancy and trophoblastic disease, as well as two frightful ones, on common denominators of pregnancy and malignancy and on the immune system in gynaecological malignancy. The last 95 pages, however, on fertility immunology, would have merited publishing on their own. All nine chapters are worth reading, particularly Bigazzi's excellent review of the immunological effects of vasectomy.

There are virtually no new data in this book. Out of 32 chapters, twelve are fair to good, nine bad to ghastly and the rest average. Reading it is likely to make the obscurity of pregnancy immunobiology darker yet, at a time when the relative importance of syncytiotrophoblast antigens, the maternal responses to them,

and the basic genetics of human pregnancy and its diseases are beginning to make a discernible pattern. Half of the contributors come from New York, which may be why some chapters have an insular feel — the authors are seemingly unaware of work going on elsewhere in the world. I

cannot imagine that anyone would want to buy this book for themselves, as all the information it contains is already available in cheaper and better presented forms. □

Colin Stern is a Consultant in Paediatrics at St Thomas' Hospital, London.

Lymphokines for the nineteen eighties

Dudley C. Dumonde

The Lymphokines: Biochemistry and Biological Activity. Edited by John W. Hadden and William E. Stewart II. Pp.437. ISBN 0-89603-012-1. (Humana/Wiley: 1981.) \$59.50 in the United States, \$69.50 elsewhere.

IN 1982 the serious student of immunology need no longer doubt the existence of lymphokines: he has only to consult the literature. As non-antibody protein products of lymphocyte activation, lymphokines act as mediators and regulators of many aspects of lymphocyte function. After a slow start in the 1960s the pace of lymphokine research began to quicken. Semantic barriers were broken down: between mouse-doctors studying "help"; guinea-pig doctors studying "delayed-type hypersensitivity"; and human doctors studying "disease". By the end of the 1970s two international conferences on lymphokines had been held, a special review journal devoted to the subject had been launched, and lymphokine research had become one of the most vigorous fields of molecular and cellular immunology.

Now in 1982, gone are the days when sceptics viewed the term "lymphokine" as having merit only in so far as it could be used both as a noun and as an adjective. The application to lymphokine research of brighter biochemistry, supplemented by techniques of cell hybridization, cell cloning and gene cloning, has even awakened venture capital: not only is there a ferment of lymphokine literature but also a ferment of medical interest in the diagnostic and therapeutic potential of lymphokines. Accordingly the lymphokinologist of the 1980s will find that Dr Hadden and Dr Stewart have edited a book which must be compulsive reading and which must be compulsively purchased given the current pressures of information retrieval.

In *The Lymphokines* the discerning consumer will find his favourite item delicately positioned between an appetizing prologue by Byron Waksman and a satisfying epilogue by Barry Bloom. Considerable effort has been made by the contributors — all authorities — to emphasize what is known about the physico-chemical nature of lymphokines and the biochemical events that attend their production and action. This theme is well

illustrated in the first half of the book by chapters on chemotactic, migration-inhibitory, macrophage-activating, cytotoxic, thymocyte-stimulating and colony-stimulating lymphokines. A most useful feature is the inclusion of two chapters on the confusing subject of helper and suppressor factors (though at the time the book was finally edited, the term "interleukin" would seem to have only just been coined). There is a chapter on interferons by Dr Stewart himself; here I would like to have seen a more critical discussion of the relationships between interferons and lymphokines.

Having dealt with what we all regard as "lymphokines", the second half of the book is comprised of chapters on leukocyte transfer factor, immune RNA, macrophage secretion products including prostaglandins, monocyte/granulocyte haemopoiesis and thymus hormones. To this extent the book is heterogeneous, though the student of lymphokines cannot afford to ignore current knowledge of other factors apparently regulating lymphocyte and leukocyte function. The general standard of production is satisfactory, though the subject index lacks details and there are quite a number of typographical errors.

Yes! This is a useful addition to the lymphokine literature, although it is not necessarily the "key resource" on the subject. If the aim of the book is to relate lymphokine biochemistry to biological activity, then it must be regarded as premature, simply because much of this research remains to be done. In fact, it seems to be this very theme that runs through many of the contributions, as if the authors were aware that whilst they were writing their reviews someone else was busy cloning the mRNA of interleukin 2 in *Xenopus* oocytes, or copying the gene for γ -interferon in *E. coli*. Can one capture at any moment the pace of a fast-moving scientific field with all of its genetic, biochemical, biological and clinical implications? Drs Hadden and Stewart, together with their well-known contributors, have sallied forth where some others may have feared to tread. The result is appealing. □

Dudley C. Dumonde is Professor of Immunology and Head of the Department of Immunology at St Thomas' Hospital and Medical School, London.

Deep Levels in Semiconductors

M Jaros

This educative monograph is written to meet the current need for an authoritative survey of the physics of deep-level impurity systems in semiconductor materials. An understanding of deep levels is important because they play a significant part in many processes affecting the functioning of semiconductor devices. The advent of more sophisticated devices has prompted recent research efforts from which new insights have emerged concerning the nature of deep levels. This book summarises the new consensus of opinion, which has previously been confined to the research literature, and provides a conceptual and quantitative understanding of a wide range of observed phenomena and related problems involving deep levels. It will be useful to researchers in experimental and theoretical semiconductor physics and devices, and to advanced students.

June 1982 xi + 302pp
0-85274-516-8 £24.00

Medical Physics Handbooks 12

Radionuclide Techniques in Clinical Investigation

P W Horton

A practical guide to the use of radiopharmaceuticals in clinical investigation. The author describes the production of radionuclides (using nuclear reactors, medical cyclotrons and radionuclide generators), safety aspects of their administration to patients, and the choice of which radionuclides to use for particular clinical applications. He also discusses the commonly used methods of measuring radioactivity, both *in vivo* and *in vitro*, and explains the principles of a wide range of clinical applications. Intended for advanced undergraduate and postgraduate students and research workers in nuclear medicine, technicians and radiographers with some previous training in this field, and life scientists using radioactivity in their researches.

June 1982 x + 172pp
0-85274-503-6 £13.95

Adam Hilger Ltd
Techno House
Redcliffe Way
Bristol BS1 6NX
England



Oceanography from space

John R. Apel

Spaceborne Synthetic Aperture Radar for Oceanography. Edited by Robert C. Beal, Pat S. DeLeonibus and Isadore Katz. Pp.216. ISBN 0-8018-2668-3. (The Johns Hopkins University Press: 1981.) \$29.25, £17.

THE US spacecraft Seasat, launched in June 1978, carried among its instrument complement a synthetic aperture imaging radar, SAR, operating at a radar wavelength of approximately 21 cm. During the satellite's short, three-month life, this instrument provided about 1.5 million km of imagery, mapped in a 100-km-wide swath at a resolution in excess of 40 m. The high resolution imagery represents a new and important vantage point for viewing phenomena occurring near the surface of the sea. This volume, stemming from a symposium held at The Johns Hopkins University Applied Physics Laboratory in the spring of 1980, presents the first broad, coherent publication of results and understanding of radar imagery of the ocean surface.

The SAR has provided images of the sea that often exhibit puzzling and varied features which must be interpreted in terms of changes in the small-scale roughness (near 30 cm length) of the surface. The original function of the Seasat SAR was to provide imagery of long-length surface gravity waves, from which one might derive a badly needed quantity, namely two-dimensional surface wave spectra, via Fourier transform techniques. The ocean surface as seen through the radar's eye has proved to be marvellously more complicated than anticipated, however, with such additional features as internal waves, current boundaries, eddies, wind stress variations, rainfall, oil slicks and shallow bottom topographic features appearing in the radar image. While surface waves with lengths greater than perhaps 100 m can often be seen in SAR images as periodic modulations, the functional relationship between the power spectral density obtained from the film and the surface wave psd is essentially unknown. Indeed, there is much discussion in the book of the basic hydrodynamic mechanisms that allow long waves to be imaged via variations in 30 cm wave energy. It follows that the theory of the imaging process for the ocean surface is in a relatively rudimentary state.

The book opens with generalized discussions of surface wave characteristics by Phillips and Kitaigorodskii, followed by a theoretical exposition on SAR imaging mechanisms by Harger and a review of the state of our understanding of oceanic winds by Pierson. A number of more detailed research results are presented next, classified according to winds, waves and circulation. These are punctuated by

dramatic examples of SAR images, most of which are of regions off the US east coast (thereby revealing the geographical orientation of the contributors to the book). The clear emphasis of the research papers is on geophysics, not radar technology. Then follow reproductions of some 21 full-swath SAR images of the western North Atlantic, New England and the central Atlantic states, prepared by the symposium organizer Robert Beal; these illustrate the myriad of land and ocean features that characterize Seasat radar images.

This volume represents the first organized exposition of the oceanic results

from SAR and, as such, should be available to anyone interested in this phase of the subject. Beyond this audience, however, there should be considerable interest on the part of oceanographers, marine meteorologists, and forward-looking civilian and military users of the sea. The results presented here are so unusual and compelling that the future will almost of necessity see additional imaging radars in space; however, a great deal of research work remains to be done before maximum use can be made of such instruments.

John R. Apel, now Assistant Director for Planning in the Applied Physics Laboratory at Johns Hopkins, was until recently at the Pacific Marine Environmental Laboratory of the National Oceanic and Atmospheric Administration, Seattle.

Explosion in palaeobotany

W. G. Chaloner

Paleobotany, Paleocology, and Evolution. Edited by Karl J. Niklas. Vol.1, pp.297, ISBN 0-03-059136-8; Vol.2, pp.279, ISBN 0-03-056656-8. (Praeger: 1981.) Vol.1 \$37.50, £31.25; Vol.2 \$36, £27.

PALAEOBIOLOGY has, in its own jargon, undergone a recent evolutionary explosion. Twenty years ago the study of fossil plants consisted largely of investigating their structure, and where possible reconstructing the whole plant from its variously preserved fragments. In the past two decades, however, the science has linked up with a number of bordering disciplines to encompass the development, reproductive biology, biogeography, biochemistry and a range of other aspects of past plant life. These two volumes contain a collection of papers by workers in the United States, which illustrate this diversity of new approaches. They were presented at a symposium held in Cornell University to honour Professor Harlan P. Banks on his retirement. The theme is very fitting, since over some 35 years of research and lecturing Banks has contributed uniquely to giving fossil botany an image not only of a challenging and rigorous science, but of an exciting one for students to enter.

The volumes comprise 14 articles dealing with topics ranging from Pre-Cambrian microfossils (E. S. Barghoorn) and palaeocology (A. H. Knoll), through Devonian plant structures (C. B. Beck) and Carboniferous swamp ecology (T. L. Phillips and W. A. DiMichele) to various aspects of gymnosperm (T. N. Taylor) and flowering plant evolution (G. Retallack and D. L. Dilcher; J. A. Wolfe and W. L. Crepet). Two chapters deal largely with extant plants in the context of their evolutionary history — T. Swain and G.

Cooper-Driver with biochemical strategy of primitive land plants, and R. M. Schuster with liverwort evolution. Three deal with a broader perspective of evolutionary processes, B. H. Tiffney reviewing changing diversity through the history of land plants, T. J. M. Schopf the process of genomic change, and A. M. Ziegler and others the biogeography of plants through the Palaeozoic.

Evidence for the interaction of plants and past environments is a recurrent theme. The study of microfossils is still very young, and the prospect of unravelling their ecology, novel. But Knoll is able to demonstrate convincingly the correlation between environment of deposition and various aspects (species diversity, association) of Precambrian microfossils. The occurrence of this "horizontal" variation between different assemblages, in addition to the well-documented "vertical" changes, is important for two aspects of Precambrian work: if such microfossils are to be used in dating rocks, the environmental control of assemblages must be understood; equally, if we are to interpret this early phase of plant history, it is important not to confuse migration in response to environmental change with evolutionary innovation. A different aspect of palaeocology is reviewed by Phillips and DiMichele in a quantitative review of the vegetation of the Carboniferous coal swamps, preserved in a three-dimensional state in coal balls of the American Mid-West. These give quantitative data on biomass, species diversity and succession. Remarkably few genera constitute the dominant plants in the assemblages, in contrast to present-day humid tropical forest. Little evidence of directional succession emerges, and abiotic factors (water level, salinity) were evidently

overriding controls.

Crepet considers the reversion from insect- to wind-pollination in the catkin-bearing trees early in the history of flowering plants. This apparently occurred in the seasonally dry tropics, so that the present-day association of this group with temperate deciduous forest is misleading as a guide to its origin. Wolfe examines the Late Tertiary replacement of broad-leaved forest by conifers in the Pacific north-west. For those overwhelmed by the complex interactions of environmental factors, it is refreshing to learn that the "sole factor" in the assumption of dominance by conifers in that region was the declining summer temperature.

Perhaps the most far-reaching contribution is Schopf's reconciliation of molecular and palaeontological approaches to evolution. He concludes that "speciation may be so instantaneous geologically that many such events might well occur within the best possible resolution of the fossil record". He argues that even if our record of past life is the product of gradualistic change, the process of its formation puts its own punctuated stamp upon the picture. Which brings us, appropriately, back to Darwin's reservation over the "imperfections" of the fossil record. □

W. G. Chaloner is Professor of Botany at Bedford College, University of London.

Room for giants

Janet V. Watson

Igneous Rocks of the British Isles. Edited by D.S. Sutherland. Pp.645. ISBN 0-471-27810-6. (Wiley: 1982.) £55, \$132.

FOR more than three thousand million years, melting in the outermost parts of the Earth has been an unusual event taking place only when and where temperatures rose to abnormally high levels. Igneous rocks formed by the consolidation of silicate melts therefore appear in the geological record as indicators of anomalous disturbances which are almost always related in time and space to major structural changes in the Earth's crust and mantle.

A book designed to document the record of igneous activity in the British Isles must therefore aim to throw light on the dynamic evolution of the entire crustal province if it is to be more than a source-book. From many points of view, there is a real need for a work of this kind. As Dr Sutherland notes in her editorial preface, no general synthesis has been published since Geikie's *Ancient Volcanoes of Britain* appeared in 1897. The literature which has accumulated since that date is vast, chaotic and

often contradictory. Embedded in it are the outlines of fundamental ideas on the genesis, differentiation and emplacement of magmas which have become classics used all over the world. Encouraged as much by the exceptionally attractive dust-jacket as by the table of contents, geologists will therefore open this book with thankfulness and hope.

Much care has been taken to make it as useful as possible as a work of reference. The treatment is broadly chronological and each of the seven parts opens with a general scene-setting chapter. The distribution of igneous suites and the structure of individual complexes are recorded in many clear maps, well-designed for their purpose. The descriptive chapters conform to a fairly consistent pattern which makes for ease of reference. Chemical analyses, geochronological data and petrographical details are segregated in appendices (the petrographical appendix living up to a hallowed tradition that goes back to J.J.H. Teall). All these are excellent features, due, no doubt, to good planning and editing.

The contents of the descriptive chapters vary in scope and style according to the kind of information available and the predilections of individual authors. The Tertiary province comes off best, with a cluster of chapters covering the volcanics, the centres, the dykes and the petrogenetic story collectively building up a coherent picture in which details fall easily into place. Elsewhere, there are some unexpected gaps (why does the best-exposed British ophiolite, in Shetland, rate only a passing mention?) and a number of thin patches not all of which can be put down to paucity of evidence.

More disturbing, however, is the apparent shapelessness of the sections covering the central event of British geological history, the Caledonian orogenic cycle. The decision to deal separately with volcanic and "plutonic" assemblages bears especially hard on these sections where the progression of tectonic events through subduction to collision and uplift is almost lost sight of. Though defensible in terms of convenience, the separation leads to absurdities such as the allocation of the Glencoe ring complex and the ignimbrites erupted in this complex during cauldron subsidence not only to different chapters but even to different parts. The British Caledonides emerge from this book as untidily as they do in the field. One longs in vain for the authoritative touch that might have shaken this sprawling orogen into shape. The Tertiary province has never lacked giants who measured up to its challenges. There is still, on the evidence of this book, something left for giants to do in the older parts of Britain. □

Janet Watson is a Professor in the Department of Geology at Imperial College, University of London.

Peat and prehistory

A. G. Smith

The Archives of the Peat Bogs. By Sir Harry Godwin. Pp.229. ISBN 0-521-23784-X. (Cambridge University Press: 1981.) £25, \$49.50.

IT HAS probably come as a surprise to many a young scientist to realize that scientific work is as much about people as it is about science. This is amply illustrated in Sir Harry Godwin's book. It is also an adventure story. Through it all runs a sense of compelling interest and excitement. Above all, however, the author illustrates the way in which ideas about the nature and importance of peat bogs have flowed down the generations of scientists. It becomes, incidentally, an interesting contribution to the history of science.

In 1935 we find Godwin confessing "my ignorance of bog ecology was vast". But he began to learn his trade in the "magical Irish countryside" (acting as chauffeur of a model T Ford) on an excursion to Ireland with Hugo Oswald and A. G. Tansley. They chanced to meet, in Roundstone, another group including Knud Jessen, the great Danish Quaternary scientist, and one of his students, G. F. Mitchell. This group had just embarked on their studies of Irish bogs. We are treated to an enduring cameo of one of the fruits of this meeting: the party slowly sinking into the mire as they listened to an apparently interminable discussion between Jessen and Oswald, in interminable Irish rain, on Von Post's concept of soligenous bog.

From this encounter came the idea among the younger workers of pooling their experience to work on an uninvestigated British peat bog. Thus in 1936 began the classical work on the stratigraphy and ecology of Tregaron Bog in Dyfed, now a National Nature Reserve, which is briefly reviewed. The story continues with Godwin's adventures in the Somerset Levels and Borth Bog (Cors Fochno), Dyfed. The first scientific investigation of this now famous bog (again a National Nature Reserve) was apparently nearly frustrated by the presence of a wide drain. Having thrown their impedimenta over it, Godwin and his companion (D. H. Valentine) had no alternative but to undress and swim for it.

It is apparent that Harry Godwin soon fell in love with British peat bogs, not only as archives, but because of their flora. This he affectionately describes, treating the various species as the *dramatis personae* of the historical pageant. In his descriptions of the bog plants he takes care to show how they can still be recognized after their preservation in the peat. These tricks of the trade are one of the many valuable side-lights of the book.

The ability to recognize the sub-fossil remains of bog plants is of course of fundamental importance in understanding

the history of the bogs themselves. The recognition of distinct layers of the remains of the giant sword sedge (*Cladium mariscus*) and their relationship to the prehistoric trackways in the peats of the Somerset Levels enabled Godwin (working initially with A. R. Clapham) to piece together their basic history. The substantial part of the book given over to this work is not only a compelling detective story: it also includes a rather charming love story for good measure! The treatment of the Somerset Levels is particularly valuable. It brings together in an easily digestible form material from a number of scattered papers. Like the work as a whole, this account can be read with equal pleasure and profit by the research student and the general reader with interests in natural history and archaeology.

The latter part of the book deals with the discovery of the effects of prehistoric farmers on the vegetation of Britain, and the impact of the new technology of radiocarbon dating on the subject. There is shorter coverage of climatic influences and sea-level changes, the origin of blanket bogs and their subsequent erosion.

Godwin's account ends intentionally at about 1960, though with most of his topics he does allude to subsequent work. How happy it is that he has been inspired to write it as a personal story. In doing so he has in no way compromised his science. The book is a pleasure to read: a book to be re-read and re-enjoyed, not only used as a work of reference. □

A. G. Smith, a former research student of Sir Harry Godwin, is Head of the Department of Plant Science at University College, Cardiff.

Drawing the line

Peter D. Moore

Wallace's Line and Plate Tectonics. Edited by T.C. Whitmore. Pp.91. ISBN 0-19-854545-2. (Oxford University Press: 1982.) £15, \$39.

IT WAS in a letter to H.E. Bates, written in 1858 from the Malay Archipelago, that Alfred Russel Wallace first drew attention to the two distinct and rigidly circumscribed faunas which are found in that region. They are as different, claimed Wallace, as the faunas of Europe and North America, yet there is no obvious geographical reason for their separation. Often the line of demarcation passes between islands lying close together.

The cause of this biogeographic discontinuity has provided much fuel for debate and the area featured strongly in the proposals of Wegener in support of his theory of continental drift. With the modern developments in plate tectonics much of the mystery has been removed from the subject, but the area remains one of the clearest demonstrations of the biogeographical consequences of moving continents. This book is a collection of essays by various authors setting forth the current state of knowledge concerning relationships between various plant and animal distribution patterns and the tectonic development of the area.

An historical background to the subject is used as an introduction and emphasis is placed on Wallace's own indecision regarding which side of the islands of Celebes his line should lie. As a result of a succession of faunal and floral analyses during the past 120 years, the line has often migrated in the minds of men. Two essays trace the continental movements leading up to the establishment of the discontinuity, and are accompanied by some detailed cartographic reconstructions. On the basis of geological data, Audley-Charles considers that the line of convergence of the Australia-New Guinea plate with Asia actually runs through the island of Celebes, thus explaining its biogeographical complexity.

The collision was a relatively recent one, occurring about $15-5 \times 10^6$ years ago, but there has still been considerable time for floral and faunal migration, mixing and even evolution, and the remainder of the book consists of a series of analyses of the biogeography of specific groups of organisms (vertebrates, palms, and other plants) in the light of recent palaeoclimatic and palaeoenvironmental reconstructions.

The papers presented in this book are concise, clear, well-illustrated reviews of a variety of specialist subjects, all of which impinge upon the modern view of Wallace's line. □

Peter D. Moore is Senior Lecturer in the Department of Plant Sciences, King's College, University of London.



Scene in New Guinea with characteristic animals (from the *Geographical Distribution of Animals*, 1876), showing Wallace's selection of Papuan animals to contrast with the Sundaic animals. Left to right and top to bottom: a tree kangaroo *Dendrolagus inustus*, the fairy lory *Charmosyna papou*, the twelve-wired bird of paradise *Seleucid melanoleuca*, the common paradise kingfisher *Tanysiptera galatea*, and a crowned pigeon *Goura cristata* (*coronata*). In the companion plate to illustrate the Sundaic fauna, Wallace chose the western tarsier, flying lemur, pentail treeshrew, Malay tapir and lesser mousedeer.